BMJ Open

Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa, and Uganda

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025469
Article Type:	Protocol
Date Submitted by the Author:	22-Jul-2018
Complete List of Authors:	Stevenson, Anne; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Akena, Dickens; Makerere University, College of Health Sciences, School of Medicine, Department of Psychiatry Stroud, Rocky; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Atwoli, Lukoye; Moi University College of Health Sciences, School of Medicine, Department of Mental Health Campbell, Megan; University of Cape Town, Department of Psychiatry and Mental Health Chibnik, Lori; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Kwobah, Edith; Moi Teaching and Referral Hospital, Department of Mental Health Kariuki, Symon; KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medical Research – (Coast); Oxford University, Department of Psychiatry Martin, Alicia; Broad Institute, Stanley Center for Psychiatric Research; Massachusetts General Hospital, Analytic and Translational Genetics Unit de Menil, Victoria; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Newton, Charles; KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medical Research – (Coast); Oxford University, Department of Psychiatry Sibeko, Goodman; University of Cape Town, Department of Psychiatry and Mental Health Stein, Dan; University of Cape Town, Department of Psychiatry and Mental Health; University of Cape Town, SA MRC Unit on Risk & Resilience in Mental Disorders Teferra, Solomon; Addis Ababa University College of Health Sciences, School of Medicine, Department of Psychiatry Zingela, Zukiswa; Walter Sisulu University, Department of Psychiatry and Human Behavioral Sciences Koenen, Karestan; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research
Keywords:	GENETICS, MENTAL HEALTH, Africa, Schizophrenia & psychotic disorders < PSYCHIATRY

Note from the Editors: Instructions for reviewers of study protocols

Since launching in 2011, BMJ Open has published study protocols for planned or ongoing research studies. If data collection is complete, we will not consider the manuscript.

Publishing study protocols enables researchers and funding bodies to stay up to date in their fields by providing exposure to research activity that may not otherwise be widely publicised. This can help prevent unnecessary duplication of work and will hopefully enable collaboration. Publishing protocols in full also makes available more information than is currently required by trial registries and increases transparency, making it easier for others (editors, reviewers and readers) to see and understand any deviations from the protocol that occur during the conduct of the study.

The scientific integrity and the credibility of the study data depend substantially on the study design and methodology, which is why the study protocol requires a thorough peer-review.

BMJ Open will consider for publication protocols for any study design, including observational studies and systematic reviews.

Some things to keep in mind when reviewing the study protocol:

- Protocol papers should report planned or ongoing studies. The dates of the study should be included in the manuscript.
- Unfortunately we are unable to customize the reviewer report form for study protocols. As such, some of the items (i.e., those pertaining to results) on the form should be scored as Not Applicable (N/A).
- While some baseline data can be presented, there should be no results or conclusions present in the study protocol.
- For studies that are ongoing, it is generally the case that very few changes can be made to the methodology. As such, requests for revisions are generally clarifications for the rationale or details relating to the methods. If there is a major flaw in the study that would prevent a sound interpretation of the data, we would expect the study protocol to be rejected.

Title: Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa, and Uganda

Author List: Anne Stevenson[1,2*], Dickens Akena[3*], Rocky E. Stroud[1,2], Lukoye Atwoli[4], Megan M. Campbell[5], Lori B. Chibnik[1,2], Edith Kwobah[6], Symon M. Kariuki[7,8], Alicia R. Martin[2,9], Victoria de Menil[1,2], Charles R.J.C Newton[7,8], Goodman Sibeko[5], Dan J. Stein[5,11], Solomon Teferra[12], Zukiswa Zingela[13], Karestan C. Koenen[1,2]

Affiliations:
[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA
[2] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, USA
[3] Department of Psychiatry, School of Medicine, College of Health Sciences, Makerere University, Kampala, Uganda
[4] Department of Mental Health, School of Medicine, Moi University College of Health Sciences, Eldoret, Kenya
[5] Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa
[6] Department of Mental Health, Moi Teaching and Referral Hospital, Eldoret, Kenya
[7] KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
[8] Department of Psychiatry, University of Oxford, Oxford, UK
[9] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[10] SA MRC Unit on Risk & Resilience in Mental Disorders, University of Cape Town, Cape Town, South Africa
[11] Department of Psychiatry, School of Medicine, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia
[12] Department of Psychiatry and Human Behavioral Sciences, Walter Sisulu University, Mthatha, South Africa

**Shared first author*

Corresponding author: Anne Stevenson
Address: Stanley Center for Psychiatric Research, Broad Institute, 75 Ames Street, 9th Floor, Cambridge, MA 02142 USA
Email: astevens@broadinstitute.org; Phone: +1 617 714 8133

Word count: 3,691 excluding title page, abstract, strengths/limitations, references, figures

ABSTRACT

Introduction: Schizophrenia and bipolar disorder account for a large proportion of the global burden of disease. Despite their enormous impact, little is known about their pathophysiology. Given the high heritability of schizophrenia and bipolar disorder, unbiased genetic studies offer the opportunity to gain insight into their neurobiology. However, advances in understanding the genetic architecture of schizophrenia and bipolar disorder have been based almost exclusively on subjects of Northern European ancestry. The Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis) project aims to expand our understanding of the causes of schizophrenia and bipolar disorder through large-scale sample collection and analyses in understudied African populations.

Methods and analysis: NeuroGAP-Psychosis is a case-control study of 34,000 participants recruited across multiple sites within Ethiopia, Kenya, South Africa, and Uganda. Participants will include individuals who are at least 18 years old with a clinical diagnosis of schizophrenia or bipolar disorder ("psychosis") or those with no history of psychosis. Research assistants will collect phenotype data and saliva for DNA extraction. Data on mental disorders, history of physical health problems, substance use, and history of past traumatic events will be collected from all participants.

DNA extraction will take place in-country, with genotyping performed at the Broad Institute. The primary analyses will include identifying major groups of participants with similar ancestry utilizing the computation-efficient program SNP weights. This will be followed by a GWAS within and across ancestry groups.

Ethics and Dissemination: All participants will be assessed for capacity to consent using the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC). Those demonstrating capacity to consent will be required to provide informed consent. Ethical clearances to conduct this study have been obtained from all participating sites. Findings from this study will be disseminated in publications and shared with controlled access public databases, such as dbGaP.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Strength: This will be the first psychiatric genetics study at this scale in Africa and across different African populations.
- Strength: Significant phenotypic data will be collected with the goal of a looking at a transdiagnostic category, without necessarily being limited to Diagnostic and Statistical Manual of Mental Disorders (DSM)/International Statistical Classification of Diseases and Related Health Problems (ICD) diagnoses.
- Strength/limitation: We will extract DNA from saliva samples which are logistically easier to manage than blood. However, there is less DNA in saliva than in blood and there is a higher risk of contamination.
- Limitation: As we have chosen not to exclude based on race, some of the participants are likely to be of European descent.
- Limitation: Although we obtain permission to follow-up enrolled participants, this is a case-control study and thus will only have phenotypic data at one point in time.

INTRODUCTION: Neuropsychiatric disorders are the leading cause of years lived with disability in the world.[1] Within mental, neurological, and substance use disorders, schizophrenia and bipolar disorder account for more than 14% of years of life lost to premature mortality and years lived with disability.[2] In the past several years, there have been great strides in our understanding of the genetic architecture of schizophrenia and bipolar disorder. A landmark paper in *Nature* from 2014 discovered 108 genome-wide significant loci for schizophrenia in ~37,000 cases and ~113,000 controls.[3] In the intervening years, the number of genome-wide significant hits for schizophrenia has grown to 145.[4] Successes in schizophrenia research have demonstrated that extremely large scale meta-analyses are necessary to identify genetic variants associated with neuropsychiatric disorders.[5] The hope is that these breakthroughs in neuropsychiatric genetics will lead to new pharmacological targets and ultimately treatments to reduce the global burden of psychiatric disorders.

However, for historical, cultural, financial, and practical reasons, these genetic findings are based predominantly on subjects of Northern European ancestry, with a growing but still small proportion on populations of East Asian ancestry.[6 7] Currently, there are major limitations in our knowledge of the genetic and environmental risk architecture of psychiatric disorders in persons of African descent.[8] As a result, we are limited in our ability to understand biological mechanisms, predict genetic risk[9], and produce optimal therapy for African populations. Moreover, African genomes are characterized by shorter haplotype blocks and contain almost a million more variants per individual than populations outside Africa.[10] Further, genetic studies of underrepresented populations afford the opportunity to discover novel loci that are invariant in European populations[11]. Thus, including data from African populations in genetic studies of neuropsychiatric disorders may accelerate genetic discovery and could be useful for fine mapping of disease causing alleles.[12] Studies of psychiatric genetics are in their infancy in Africa and are not yet at a scale necessary for variant discovery.

AIM AND OBJECTIVE: The Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis) project aims to expand knowledge of the genetic and environmental risk factors for neuropsychiatric disorders in Africa through large-scale sample collection and analysis, so that future advances in science and therapeutics can account for and be applicable to African populations.

METHODS AND ANALYSIS:

Study design: The design will be a case-control study. This project is structured around two diagnostic categories: schizophrenia and bipolar I disorder (grouped under the heading "psychotic disorders"). The rationale for grouping schizophrenia and bipolar disorder stems from literature showing a high level of genetic correlation between schizophrenia and bipolar disorder, which may indicate that some of the same genetic variants confer risk for both phenotypes.[13]

Cases will be individuals with a diagnosis of schizophrenia or bipolar disorder, referred to subsequently as psychosis. Controls will be individuals from the same geographic location, without psychosis, who will be matched to cases for age, sex, and ancestry.

Study sites: The study will be conducted over four years, starting in September 2018 and ending in 2022, at several sites in Ethiopia, Kenya, South Africa, and Uganda (Figure 1)[14]. DNA extraction will be performed on site and genotyping will performed at the Broad Institute in the USA using the Illumina Global Screening Array. Sites were selected on the basis of the following criteria: 1) proven track record of psychiatric research; 2) availability of research personnel and the necessary research infrastructure to be able to recruit thousands of participants; 3) existing trusted relationships from prior collaborations. Each of the countries where participants will be recruited from has enormous genetic diversity within and between them, which is likely to improve the ability of this research to answer the study objectives. Pilot studies are currently underway in all four countries.

Ethiopia. Participants will be recruited from the Amanuel Mental Specialized Hospital (cases) and Black Lion Hospital also known as the Tikur Anbessa Hospital (controls).

Kenya. In Kenya, participants will be recruited from the Moi Teaching and Referral Hospital and affiliated sites in Webuye, Kapenguria, Kitale, Kapsabet, Iten, and Kakamega and the KEMRI-Wellcome Trust Research Programme with recruiting sites in Kilifi County, Malindi sub-County, Port Reitz, and Coast General Provincial Hospitals.

South Africa. Participants will be recruited from the Western Cape, Eastern Cape and the Gauteng provinces. In the Western Cape, participants will be recruited from the Valkenberg, Lentegeur, Khayelitsha District, and Groote Schuur Hospitals as well as a number of community clinics. In the Eastern Cape, participants (cases) will be recruited from the Fort England Psychiatric, Elizabeth Donkin, Tower Psychiatric, and Komani Hospitals while controls will be recruited from the Nelson Mandela Academic and Dora Nginza Hospitals as well as five affiliated health clinics.

Uganda. Study participants in Uganda will be recruited from the Butabika National Mental Health Referral Hospital (cases and controls), Naguru (controls only), Arua (cases and controls), Mbarara (cases and controls), and Gulu Regional Referral Hospitals (cases and controls).

Inclusion criteria: Individuals with a clinical diagnosis of psychosis (cases) as confirmed by clinician referral and/or medical record review, and those without a clinical diagnosis of psychosis (controls) will be eligible to participate. All participants (cases and controls) will be required to provide written informed consent or a fingerprint in case of illiteracy and must be at least 18 years old. To ensure that participants (cases and controls) have sufficient capacity and autonomy to consent to the study, we will use the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC),[15] which has been used as an iterative learning tool in similar populations in South Africa.[16]

Exclusion criteria for cases: Individuals (cases) will be excluded if the following are present:

- Absence of a diagnosis of a psychotic disorder
- Severe, intrusive levels of psychiatric symptoms at the time of consent
- Intoxication or withdrawal from of alcohol or substance abuse at the time of consent
- A current psychiatric hospitalization (inpatients)
- Involuntary detention at the time of consent
- Lack fluency in one of the languages the consent form has been translated into
- Lack capacity to consent to the study as determined by the UBACC.

Potential cases may be approached at a later date once their symptoms are controlled if they meet other inclusion criteria. There will be no exclusions for cases or controls based on sex, ancestry/ethnicity, religious affiliations, or sexual orientation.

Exclusion criteria for controls: Potential controls will be excluded if they:

- Have current psychotic symptoms or a past diagnosis of a psychotic disorder
- Are currently taking medication for psychosis
- Have acute levels of alcohol or substance misuse as demonstrated by being a current inpatient or under acute medical care for substance misuse
- Are not fluent in any of the languages the consent form has been translated into
- Lack capacity to consent to the study as determined by the UBACC.

Participant recruitment: Eligible participants (cases) will be identified by clinical staff through review of their medical records. A research assistant (RA), who is a nurse, clinical officer, clinical assistant, or bachelor's level accredited RA and who has received study specific training and human subjects training, will approach the prospective participant, carefully explaining to them in the local language the purpose and procedures of the study and emphasizing that participation is entirely voluntary and will not impact any medical care they receive. Languages include: Acholi, Afrikaans, Amharic, English, Kigiryama, Kiswahili, Luganda, Lugbara, Runyankole, and isiXhosa.

After reading the information sheet out loud to the subject, the RA will administer the UBACC, a 10-item questionnaire that evaluates the potential participant's understanding of different components of the study. Each response will be scored on a range of 0-2, with 0 representing no understanding and 2 representing a clear understanding. In the NeuroGAP-Psychosis study, the RA will administer the UBACC over a maximum of four trials. The RA will re-explain and re-administer any items the subject answered incorrectly. The process will end if the full score of 20 is obtained. After the 4th trial, participants who are unable to achieve a score of 14.5, the cut-off originally developed for screening decisional capacity using the UBACC,[15] will be excluded.

The RA will proceed to obtain informed consent from those who express interest in participation, and ask participants to provide saliva into an Oragene tube, from which their DNA will be extracted. To maximize the quality of the collected DNA, participants will be asked not to eat, drink, smoke, or chew gum for 30 minutes prior to providing the sample during the consenting process.

Phenotypic assessments were selected on the basis of the following criteria: 1) the particular domain assessed; 2) cross-cultural validity; 3) non-proprietary measures, when possible; 4) investigators' prior experience using the tool; and 5) the time length of administering the tool. The study will collect the following battery of instruments (see Figure 2):

- Mini International Neuropsychiatric Interview, Standard 7.0.2 for DSM-5 (MINI): modules A, C, K, and O on major depressive episodes, manic and hypomanic episodes, and psychotic disorders and mood disorders with psychotic features, respectively[17]
- Life Events Checklist for DSM-5 (LEC-5): a 17-item scale covering exposures to potentially traumatic events[18]
- Alcohol, Smoking and Substance Involvement Screening Test, version 3.0 (ASSIST): a subset of the ASSIST on substance type and use over a participant's lifetime and over the past three months[19]

- Composite International Diagnostic Interview screener (CIDI): a checklist within the CIDI on chronic physical conditions[20]

Participant phenotypic data will be collected using encrypted tablets, and uploaded to a secure cloud-based server. Other sources of data will include chart reviews to ascertain clinical diagnoses and medication use by the participants; as well as measurements of blood pressure, heart rate, height, and weight.

Recruitment of controls will take place from general medical facilities (some of these facilities are located within the premises where cases will be identified). In one site, at the KEMRI-Wellcome Trust Research Programme in Kenya, controls may also be recruited from homes through the Kilifi Health and Demographic Surveillance System (KHDSS).[21] [In KHDSS, census data is recorded every four months through household surveys in the catchment area and linked with healthcare data from Kilifi District Hospital.]

The study visit for controls will match the process for cases except for two phenotypic batteries, the K10 and the PSQ (see Figure 3); instead of using the MINI for controls, the RA will administer:

- Kessler Psychological Distress Scale (K10): a 10-item questionnaire to measure anxiety and depression[22]
- Psychosis Screening Questionnaire (PSQ): a brief screening tool for psychotic conditions[23]

All participants will receive compensation to offset the time and inconvenience of participating in the study. The estimated length of a study visit is 90 to 120 minutes.

Patient and Public Involvement

Patients' priorities and experiences were taken into account in the design of the study. Prior research and clinical work with people with neuropsychiatric disorders at each of the sites informed the research question and study design, including limiting the length of the study battery, minimizing intrusive sample collection (saliva instead of blood), and incorporating a tool as part of the consent process to assess understanding of the research and study visit. In addition, most of the sites have a precedent of genetics research and we engaged with teams from those other genetic studies to incorporate patient concerns emerging from genetic research. Study findings will be shared with local community advisory boards.

Sample size and power calculation: Over a period of four years, the study team expects to recruit a total of 17,000 schizophrenia and bipolar disorder cases and 17,000 controls across four countries: Ethiopia, Kenya, South Africa, and Uganda. The geographical and demographic sampling distribution of the controls will follow that of the cases, and cases and controls will be matched for ethnic ancestry in all analyses. These numbers, described in more detail below, will give adequate statistical power to independently examine the genetic architecture of schizophrenia and bipolar disorder in African populations, the distribution and influence of rare copy number variation on neuropsychiatric disease risk in African populations, and determine the consistency of genetic influences on neuropsychiatric disease risk between populations of European and African descent. Ultimately these data will be analyzed in combination with other studies of psychiatric genetics, most notably from the Psychiatric Genetics Consortium, which will further increase our power to discover new and validate current variants with small effect sizes.

With 17,000 cases and controls, using a threshold of 2.5 x 10^{-8} for genome-wide significance, we have >90% power to find common variants (Minor Allele Frequency (MAF) >5%) with effects of 1.15 or greater.[24] For more common variants with MAF=25%, we have 89% power to see significant effects as low as 1.06. These effect sizes are within the range of what was found in a recent GWAS of schizophrenia in a European population.[3]

DNA extraction and processing: All study participants will provide a 2.0 mL sample of saliva by spitting into a funnel in an Oragene(R) OR-500 kit. DNA extraction will take place in-country, with one aliquot of the extract sent to the Broad Institute in Cambridge, MA, USA for genotyping. The remaining DNA will be frozen and stored in-country within designated biobanks and will be governed by the rules of its institution. In Ethiopia, DNA will be stored at the Black Lion Hospital. In Kenya, DNA will be stored at Moi University and KEMRI-Wellcome Trust Research Programme Biorepository Laboratories. In South Africa, DNA will be stored in a biobank at the Division of Human Genetics at the University of Cape Town, while in Uganda, DNA will be stored at the Integrated Biorepository Lab of H3Africa in Uganda (IBR-H3A-U).

All samples will be assayed on the same genotyping array at the same facility. To maximize information and allow for comparisons across studies genotyped on different platforms, we will impute SNP genotypes not present on specific arrays. Genotype imputation of untyped markers will be performed using default parameters in IMPUTE2 using appropriate reference panels.

Training: To ensure research staff understand the study and their responsibilities, teams will undergo multiple trainings. Study staff who have contact with research participants will complete the online Human Subject Training offered by the US National Institutes of Health. Site leadership and Harvard staff will provide in-person training to the teams by providing an overview of the project, leading discussions, providing role playing on ethics and confidentiality, recruiting and consenting participants, on reportable events, and on storing forms and data properly. Additionally, there will be hands-on training using the electronic tablets, collecting saliva using Oragene kits, administering the phenotypic batteries, and taking physiological measures, such as height and weight. A checklist will be created of key skills staff need before they can begin interacting with participants one-on-one (Figure 4). This form will need to be signed and dated by the staff member and the staff member's supervisor before he or she can begin working with participants; it will then be uploaded to the regulatory binder. Harvard teams will reinforce these skills with multiple site visits per year and weekly video calls.

In addition, Broad Institute wet lab staff will work in partnership with the lab teams in Africa to standardize the DNA extraction process across the five sites and to increase the quality of the extractions through visits, video calls, and email. Where needed, the Broad staff will train lab members on DNA extraction from saliva and establish a sample chain of custody process in order to decrease the likelihood of sample swaps.

Statistical analysis plans:

a) Ancestry analyses: In contrast to most samples previously analyzed by the psychiatric genetics community – which were predominately of European ancestry[25] – the NeuroGAP-Psychosis samples will include individuals of predominantly African ancestry and will be drawn from multiple populations across East and South Africa where genetic variation will be more structured and demographically complex[26 27]. Therefore, before proceeding with any phenotype analyses we will first analyze genetic ancestry in our cohorts to identify large, homogeneous groups to conserve optimal power in the GWAS and ultimately control for additional population stratification within these groups to avoid spurious results.

We will take advantage of the large amount of genotypic information available and control for potential population stratification in a standard and validated two-step process. First, we will compute principal components adjusting for population substructure and relatedness using PC-AiR[28]. We will define clusters of ancestries using ADMIXTURE[29], and choose the optimal number of clusters with 5-fold cross-validation to identify major groups of participants with similar ancestry. The inclusion of extensive population reference samples compiled by our group and available from other studies will increase the power of this approach and lead to stable individual ancestry calling across studies. Second, if significant differences are found between populations in our cohorts, as expected, our initial analyses will then be conducted separately on more homogeneous ancestry groups, and later meta-analyzed. Within each group, we will conduct principal component analysis (PCA) again using PC-AiR, and the principle components (PCs) that explain the majority of the variation in ancestry within the cohort will be included as covariates in the regression analyses. Again adjusting for population structure, we will compute a genetic relatedness matrix using PC-Relate[30], which can be used as a random effect in a linear or logistic mixed model. In addition, as new methods are developed by our group and others to account for population structure, we will modify our analyses when appropriate.

b) Genome-wide Association Study (GWAS). The primary GWAS discovery analysis will be conducted for the dichotomous schizophrenia/bipolar ("psychotic disorders") diagnosis using logistic regression. All analyses will be performed using PLINK[31] and R.[32] To assess significance thresholds and correct for multiple comparisons, we will apply the conventional threshold of $p<2.5x10^{-8}$.

c) Heritability estimates. Estimation of heritability, the fraction of disease risk attributable to inherited genetic differences, is an important initial analytic goal. Heritability estimates speak to the anticipated architecture and power of a genome-wide association study, a primary endpoint of the NeuroGAP-Psychosis project. We will use Genome-wide Complex Trait Analysis (GCTA) to estimate SNP heritability within Africa.[25 33] We will assess genetic correlation with existing GWAS in other globally diverse populations using POPCORN[34].

ETHICS AND DISSEMINATION:

Ethics: Ethical and safety considerations will be taken across multiple levels. Since the subjects the study aims to recruit are deemed vulnerable populations, additional measures will be taken to protect them. Potential participants will be excluded if they are presenting with severe, intrusive levels at the time of consent. In addition, as described previously, the RAs will use the UBACC [15 16] during the consent process to make sure participants understand the study, what is required of them, and that they can withdraw at any point. Participants who pass the UBACC and who want to continue will be required to provide written informed consent or a fingerprint in lieu of a signature.

All steps will be taken to mitigate the potential loss of confidentiality for participants. A two-part ID system will be instituted in order to limit the possibility of connecting the participant with their data. There will be one ID for the phenotypic information, and one for genetic information. A single encrypted database linking the phenotypic ID and genetic ID will be stored in-country. The participant's name and contact information will be on the consent form, which will be stored in a locked cabinet. This information will be put in a separate secure database and will never leave the country of origin. No data that is considered identifiable, by Health Insurance Portability and Accountability Act (HIPPA) standards, will ever be collected on the electronic tablets.

Ethical clearances to conduct this study have been obtained from all participating sites, including:

- Ethiopia: Addis Ababa University College of Health Sciences (#014/17/Psy) and the Ministry of Science and Technology National Research Ethics Review Committee (#3.10/14/2018).
- Kenya: Moi University School of Medicine Institutional Research and Ethics Committee (#IREC/2016/145, approval number: IREC 1727), Kenya National Council of Science and Technology (#NACOSTI/P/17/56302/19576) KEMRI Centre Scientific Committee (CSC# KEMRI/CGMRC/CSC/070/2016), KEMRI Scientific and Ethics Review Unit (SERU# KEMRI/SERU/CGMR-C/070/3575)
- South Africa: The University of Cape Town Human Research Ethics Committee (#466/2016) and Walter Sisulu University Research and Ethics Committee (# 051/2016)
- Uganda: The Makerere University School of Medicine Research and Ethics Committee (SOMREC #REC REF 2016-057) and the Uganda National Council for Science and Technology (UNCST #HS14ES)
- USA: The Harvard T.H. Chan School of Public Health (#IRB17-0822)

Dissemination: We aim to share results through a number of mechanisms in addition to conferences and peer-reviewed publications. In order to maximize scientific utility of the samples/data and to minimize data waste, both genomic and phenotypic data from NeuroGAP-Psychosis will be deposited in controlled access public databases, such as the database of Genotypes and Phenotypes (dbGaP), the Psychiatric Genomics Consortium (PGC), and/or the European Genome-Phenome Archive (EGA). By pooling NeuroGAP-Psychosis' data with that from other studies, it will be possible to maximize statistical power. Sharing the data will also contribute to the "genomic revolution" currently underway in Africa due to the pioneering work of the African Genome Variation Project, MalariaGEN, African Society of Human Genetics, and Human Hereditary and Health in Africa (H3Africa).[27 35-38]

To help ensure all collaborators have the capacity to work with all the data that will be created, a unique training program has been created, the Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER),[39] that will run in parallel to the NeuroGAP-Psychosis study. Over a two-year period, 17 early-career researchers from Ethiopia, Kenya, South Africa, and Uganda will take part in online and in-person trainings on topics including biostatistics, genetic analysis, and epidemiology. Ultimately, the goal of GINGER is to build the next generation of neuropsychiatric geneticists who will be able to work on the NeuroGAP-Psychosis data and continue its legacy in the future.

Authors' contributions:
KK and DS conceptualized and designed the study.
AS, DA, RS, LA, EK, SK, VdM, CN, DS, ST, ZZ, and KK wrote the protocol, selected the phenotype questionnaires, and added site-specific information.
LC and AM wrote the sample size, power calculations, and statistical analysis plans.
MC and GS contributed to the UBACC process.
AS and DA wrote the draft of the manuscript and incorporated the revisions by the co-authors; they contributed equally to this paper.
All authors reviewed the manuscript for intellectual content, contributed to revisions, and approved the final version for publication.

Funding statement: This study is funded by the Stanley Center for Psychiatric Research at Broad Institute. The sponsor was involved in the study design, in the writing of the report, and in the decision to submit the paper for publication. The sponsor will be involved in the analysis and interpretation of the data.

Competing interests statement. Dan Stein has received research grants and/or consultancy honoraria from Biocodex, Lundbeck, Servier, and Sun.

Legends for Figures:

Figure 1: Proposed collection sites for NeuroGAP-Psychosis

Figure 2: Phenotyping tools for study participants

Figure 3: The study process for cases and controls

Figure 4: NeuroGAP-Psychosis Training Checklist

REFERENCES:

1. Whiteford HA, Ferrari AJ, Degenhardt L, et al. Global Burden of Mental, Neurological, and Substance Use Disorders: An Analysis from the Global Burden of Disease Study 2010. In: Patel V, Chisholm D, Dua T, et al., eds. Mental, Neurological, and Substance Use Disorders: Disease Control Priorities, Third Edition (Volume 4). Washington (DC): The International Bank for Reconstruction and Development / The World Bank
(c) 2016 International Bank for Reconstruction and Development / The World Bank. 2016.
2. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet (London, England)* 2013;382(9904):1575-86. doi: 10.1016/s0140-6736(13)61611-6
3. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511:421. doi: 10.1038/nature13595 https://www.nature.com/articles/nature13595#supplementary-information
4. Pardinas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics* 2018;50(3):381-89. doi: 10.1038/s41588-018-0059-2
5. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. *Pharmacogenomics* 2009;10(2):191-201. doi: 10.2217/14622416.10.2.191
6. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016;538(7624):161-64. doi: 10.1038/538161a
7. Li Z, Chen J, Yu H, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature genetics* 2017;49(11):1576-83. doi: 10.1038/ng.3973
8. Dalvie S, Koen N, Duncan L, et al. Large Scale Genetic Research on Neuropsychiatric Disorders in African Populations is Needed. *EBioMedicine* 2015;2(10):1259-61. doi: 10.1016/j.ebiom.2015.10.002
9. Martin AR, Gignoux CR, Walters RK, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American journal of human genetics* 2017;100(4):635-49. doi: 10.1016/j.ajhg.2017.03.004
10. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015;526(7571):68-74. doi: 10.1038/nature15393
11. Wojcik G, Graff M, Nishimura KK, et al. Genetic Diversity Turns a New PAGE in Our Understanding of Complex Traits. *bioRxiv* 2017 doi: 10.1101/188094
12. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual review of genomics and human genetics* 2008;9:403-33. doi: 10.1146/annurev.genom.9.081307.164258
13. Cardno AG, Owen MJ. Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. *Schizophrenia bulletin* 2014;40(3):504-15. doi: 10.1093/schbul/sbu016
14. Ulrich T. Nations Represented in NeuroGAP-Psychosis: Broad Communications, 2018.
15. Jeste DV, Palmer BW, Appelbaum PS, et al. A new brief instrument for assessing decisional capacity for clinical research. *Archives of general psychiatry* 2007;64(8):966-74. doi: 10.1001/archpsyc.64.8.966
16. Campbell MM, Susser E, Mall S, et al. Using iterative learning to improve understanding during the informed consent process in a South African psychiatric genomics study. *PLOS ONE* 2017;12(11):e0188466. doi: 10.1371/journal.pone.0188466
17. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of clinical psychiatry* 1998;59 Suppl 20:22-33.
18. Weathers FW, Blake DD, Schnurr PP, et al. The Life Events Checklist for DSM-5 (LEC-5). *Instrument available from the National Center for PTSD www.ptsdva.gov* 2013
19. Humeniuk R, Henry-Edwards S, Ali R, et al. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): manual for use in primary care. 2010
20. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International journal of methods in psychiatric research* 2004;13(2):93-121.

21. Scott JA, Bauni E, Moisi JC, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *International journal of epidemiology* 2012;41(3):650-7. doi: 10.1093/ije/dys062
22. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. *Archives of general psychiatry* 2003;60(2):184-9.
23. Bebbington P, Nayani T. Psychosis Screening Questionnaire. *International journal of methods in psychiatric research* 1995;5(1): 11-19
24. Skol AD, Scott LJ, Abecasis GR, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature genetics* 2006;38(2):209-13. doi: 10.1038/ng1706
25. de Candia TR, Lee SH, Yang J, et al. Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. *American journal of human genetics* 2013;93(3):463-70. doi: 10.1016/j.ajhg.2013.07.007
26. Busby GBJ, Band G, Si Le Q, et al. Admixture into and within sub-Saharan Africa. *eLife* 2016;5:e15266. doi: 10.7554/eLife.15266
27. Gurdasani D, Carstensen T, Tekola-Ayele F, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* 2015;517(7534):327-32. doi: 10.1038/nature13997
28. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genetic epidemiology* 2015;39(4):276-93. doi: 10.1002/gepi.21896
29. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 2009;19(9):1655-64. doi: 10.1101/gr.094052.109
30. Conomos MP, Reiner AP, Weir BS, et al. Model-free Estimation of Recent Genetic Relatedness. *American journal of human genetics* 2016;98(1):127-48. doi: 10.1016/j.ajhg.2015.11.022
31. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007;81(3):559-75. doi: 10.1086/519795
32. R: A Language and Environment for Statistical Computing [program]. Vienna, Austria, 2018.
33. Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* 2011;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011
34. Brown BC, Ye CJ, Price AL, et al. Transethnic Genetic-Correlation Estimates from Summary Statistics. *American journal of human genetics* 2016;99(1):76-88. doi: 10.1016/j.ajhg.2016.05.001
35. Beiswanger CM, Abimiku A, Carstens N, et al. Accessing Biospecimens from the H3Africa Consortium: Biopreserv Biobank. 2017 Apr 1;15(2):95-8. doi:10.1089/bio.2017.0008.
36. Consortium THA. Enabling the genomic revolution in Africa: H3Africa is developing capacity for health-related genomics research in Africa. *Science (New York, NY)* 2014;344(6190):1346-8. doi: 10.1126/science.1251546
37. Mulder N, Abimiku Al, Adebamowo SN, et al. H3Africa: current perspectives. *Pharmacogenomics and Personalized Medicine* 2018;11:59-66. doi: 10.2147/PGPM.S141546
38. Band G, Rockett KA, Spencer CC, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* 2015;526(7572):253-7. doi: 10.1038/nature15390
39. van der Merwe C, Mwesiga EK, McGregor NW, et al. Advancing neuropsychiatric genetics training and collaboration in Africa. *The Lancet Global health* 2018;6(3):e246-e47. doi: 10.1016/s2214-109x(18)30042-1

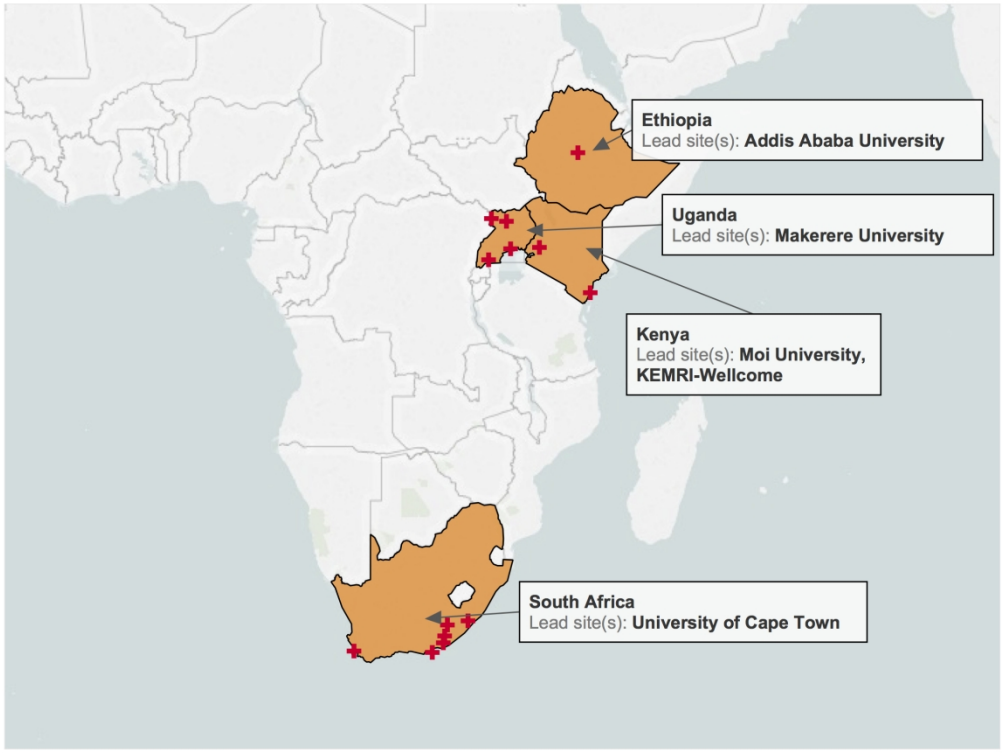

Figure 1: Proposed collection sites for NeuroGAP-Psychosis

192x145mm (300 x 300 DPI)

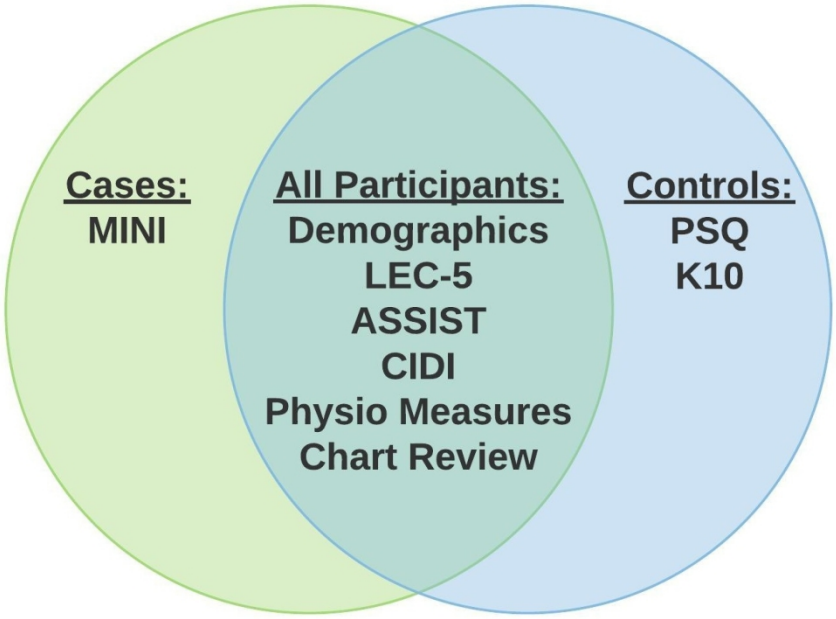

Figure 2: Phenotyping tools for study participants

130x97mm (300 x 300 DPI)

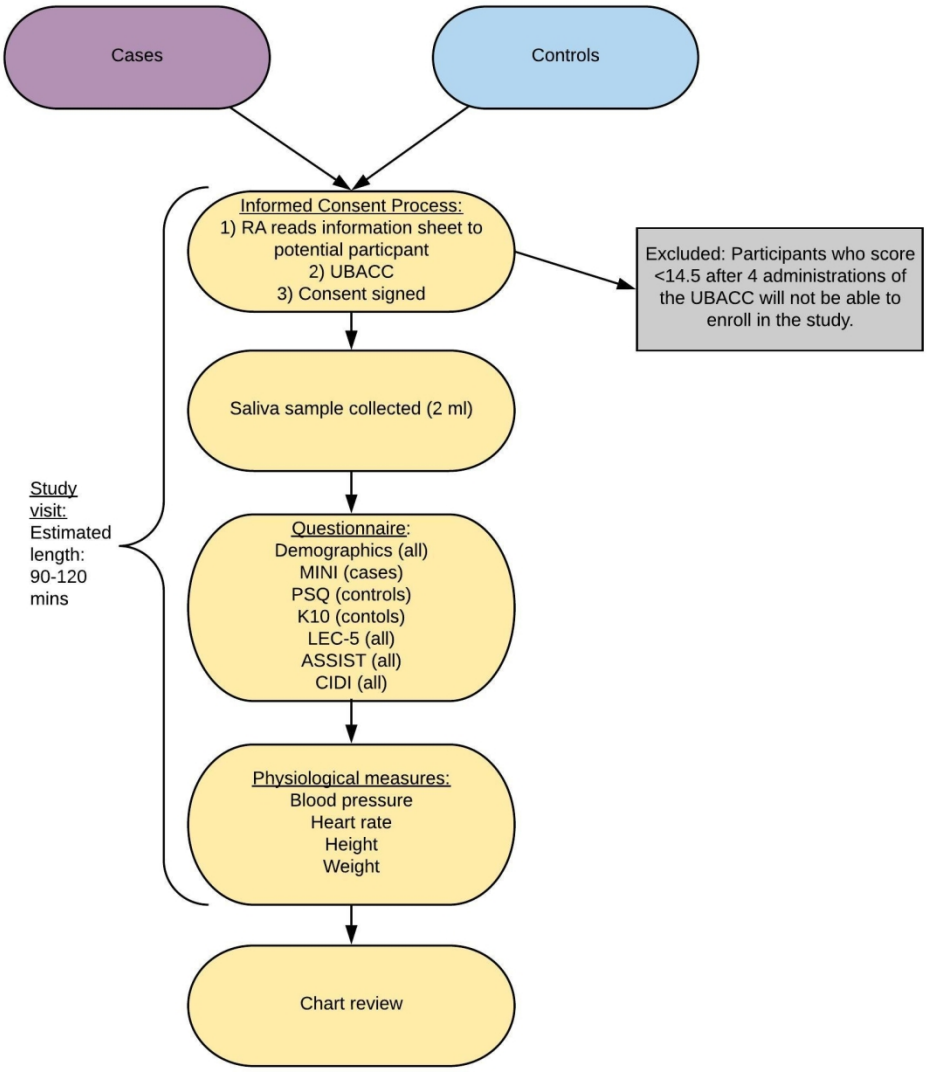

Figure 3: The study process for cases and controls

155x177mm (300 x 300 DPI)

NeuroGAP-Psychosis Training Checklist
For All Study Staff

_____
Print name above

Instructions: Check the boxes when you have completed each task. When you are done, you and your supervisor must sign and date the form to attest that you have completed the training and have approval to start working with NeuroGAP-Psychosis subjects. If some of these components are not applicable, the PI should write "Not Applicable" next to the item.

☐ Online US National Institutes of Health Ethics Training Date completed:

☐ CV on file - must be signed and dated

☐ Trained on the NeuroGAP-Psychosis Protocol

Consent Process

☐ Study Information Sheet

☐ Consent Form

☐ UBACC

Study Tools

☐ MINI ☐ LEC-5

☐ K10 ☐ PSQ

☐ ASSIST ☐ CIDI

Physiological Measures

☐ Blood pressure, heart rate, weight, height

Role Playing

☐ Role played all of the study activities from beginning to end with other members of the NeuroGAP-Psychosis team

Shadowing

☐ Watched an experienced member of the study team complete 5 study visits with participants

ATTESTATION:

_____ _____
Signature of study team member Signature of supervisor or PI

_____ _____
Print name of study team member Print name of supervisor or PI

_____ _____
Date Date

Figure 4: NeuroGAP-Psychosis Training Checklist

196x203mm (300 x 300 DPI)

BMJ Open

Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa, and Uganda

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025469.R1
Article Type:	Protocol
Date Submitted by the Author:	20-Nov-2018
Complete List of Authors:	Stevenson, Anne; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Akena, Dickens; Makerere University, College of Health Sciences, School of Medicine, Department of Psychiatry Stroud, Rocky; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Atwoli, Lukoye; Moi University College of Health Sciences, School of Medicine, Department of Mental Health Campbell, Megan; University of Cape Town, Department of Psychiatry and Mental Health Chibnik, Lori; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Kwobah, Edith; Moi Teaching and Referral Hospital, Department of Mental Health Kariuki, Symon; KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medical Research – (Coast); Oxford University, Department of Psychiatry Martin, Alicia; Broad Institute, Stanley Center for Psychiatric Research; Massachusetts General Hospital, Analytic and Translational Genetics Unit de Menil, Victoria; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research Newton, Charles; KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medical Research – (Coast); Oxford University, Department of Psychiatry Sibeko, Goodman; University of Cape Town, Department of Psychiatry and Mental Health Stein, Dan; University of Cape Town, Department of Psychiatry and Mental Health; University of Cape Town, SA MRC Unit on Risk & Resilience in Mental Disorders Teferra, Solomon; Addis Ababa University College of Health Sciences, School of Medicine, Department of Psychiatry Zingela, Zukiswa; Walter Sisulu University, Department of Psychiatry and Human Behavioral Sciences Koenen, Karestan; Harvard T.H. Chan School of Public Health, Department of Epidemiology; Broad Institute, Stanley Center for Psychiatric Research

Primary Subject Heading:	Genetics and genomics
Secondary Subject Heading:	Mental health
Keywords:	GENETICS, MENTAL HEALTH, Africa, Schizophrenia & psychotic disorders < PSYCHIATRY

Title: Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa, and Uganda

Author List: Anne Stevenson[1,2*], Dickens Akena[3*], Rocky E. Stroud[1,2], Lukoye Atwoli[4], Megan M. Campbell[5], Lori B. Chibnik[1,2], Edith Kwobah[6], Symon M. Kariuki[7,8], Alicia R. Martin[2,9], Victoria de Menil[1,2], Charles R.J.C Newton[7,8], Goodman Sibeko[5], Dan J. Stein[5,11], Solomon Teferra[12], Zukiswa Zingela[13], Karestan C. Koenen[1,2]

Affiliations:
[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA
[2] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, USA
[3] Department of Psychiatry, School of Medicine, College of Health Sciences, Makerere University, Kampala, Uganda
[4] Department of Mental Health, School of Medicine, Moi University College of Health Sciences, Eldoret, Kenya
[5] Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa
[6] Department of Mental Health, Moi Teaching and Referral Hospital, Eldoret, Kenya
[7] KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
[8] Department of Psychiatry, University of Oxford, Oxford, UK
[9] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[10] SA MRC Unit on Risk & Resilience in Mental Disorders, University of Cape Town, Cape Town, South Africa
[11] Department of Psychiatry, School of Medicine, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia
[12] Department of Psychiatry and Human Behavioral Sciences, Walter Sisulu University, Mthatha, South Africa

**Shared first author*

Corresponding author: Anne Stevenson
Address: Stanley Center for Psychiatric Research, Broad Institute, 75 Ames Street, 9th Floor, Cambridge, MA 02142 USA
Email: astevens@broadinstitute.org; Phone: +1 617 714 8133

Word count: 3,726 excluding title page, abstract, strengths/limitations, references, figures

ABSTRACT

Introduction: Schizophrenia and bipolar disorder account for a large proportion of the global burden of disease. Despite their enormous impact, little is known about their pathophysiology. Given the high heritability of schizophrenia and bipolar disorder, unbiased genetic studies offer the opportunity to gain insight into their neurobiology. However, advances in understanding the genetic architecture of schizophrenia and bipolar disorder have been based almost exclusively on subjects of Northern European ancestry. The Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis) project aims to expand our understanding of the causes of schizophrenia and bipolar disorder through large-scale sample collection and analyses in understudied African populations.

Methods and analysis: NeuroGAP-Psychosis is a case-control study of 34,000 participants recruited across multiple sites within Ethiopia, Kenya, South Africa, and Uganda. Participants will include individuals who are at least 18 years old with a clinical diagnosis of schizophrenia or bipolar disorder ("psychosis") or those with no history of psychosis. Research assistants will collect phenotype data and saliva for DNA extraction. Data on mental disorders, history of physical health problems, substance use, and history of past traumatic events will be collected from all participants.

DNA extraction will take place in-country, with genotyping performed at the Broad Institute. The primary analyses will include identifying major groups of participants with similar ancestry utilizing the computation-efficient program SNP weights. This will be followed by a GWAS within and across ancestry groups.

Ethics and Dissemination: All participants will be assessed for capacity to consent using the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC). Those demonstrating capacity to consent will be required to provide informed consent. Ethical clearances to conduct this study have been obtained from all participating sites. Findings from this study will be disseminated in publications and shared with controlled access public databases, such as dbGaP.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Strength: This will be the first psychiatric genetics study at this scale in Africa and across different African populations.
- Strength: Significant phenotypic data will be collected with the goal of a looking at a transdiagnostic category, without necessarily being limited to Diagnostic and Statistical Manual of Mental Disorders (DSM)/International Statistical Classification of Diseases and Related Health Problems (ICD) diagnoses.
- Strength/limitation: We will extract DNA from saliva samples which are logistically easier to manage than blood. However, there is less DNA in saliva than in blood and there is a higher risk of contamination.
- Limitation: As we have chosen not to exclude based on race, some of the participants are likely to be of European descent.
- Limitation: Although we obtain permission to follow-up enrolled participants, this is a case-control study and thus will only have phenotypic data at one point in time.

INTRODUCTION: Neuropsychiatric disorders are the leading cause of years lived with disability in the world.[1] Within mental, neurological, and substance use disorders, schizophrenia and bipolar disorder account for more than 14% of years of life lost to premature mortality and years lived with disability.[2] In the past several years, there have been great strides in our understanding of the genetic architecture of schizophrenia and bipolar disorder. A landmark paper in *Nature* from 2014 discovered 108 genome-wide significant loci for schizophrenia in ~37,000 cases and ~113,000 controls.[3] In the intervening years, the number of genome-wide significant hits for schizophrenia has grown to 145.[4] Successes in schizophrenia research have demonstrated that extremely large scale meta-analyses are necessary to identify genetic variants associated with neuropsychiatric disorders.[5] The hope is that these breakthroughs in neuropsychiatric genetics will lead to new pharmacological targets and ultimately treatments to reduce the global burden of psychiatric disorders.

However, for historical, cultural, financial, and practical reasons, these genetic findings are based predominantly on subjects of Northern European ancestry, with a growing but still small proportion on populations of East Asian ancestry.[6 7] Currently, there are major limitations in our knowledge of the genetic and environmental risk architecture of psychiatric disorders in persons of African descent.[8] As a result, we are limited in our ability to understand biological mechanisms, predict genetic risk,[9] and produce optimal therapy for African populations. Moreover, African genomes are characterized by shorter haplotype blocks and contain almost a million more variants per individual than populations outside Africa.[10] Further, genetic studies of underrepresented populations afford the opportunity to discover novel loci that are invariant in European populations.[11] Thus, including data from African populations in genetic studies of neuropsychiatric disorders may accelerate genetic discovery and could be useful for fine mapping of disease causing alleles.[12] Studies of psychiatric genetics are in their infancy in Africa and are not yet at a scale necessary for variant discovery.

AIM AND OBJECTIVE: The Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis) project aims to expand knowledge of the genetic and environmental risk factors for neuropsychiatric disorders in Africa through large-scale sample collection and analysis, so that future advances in science and therapeutics can account for and be applicable to African populations.

METHODS AND ANALYSIS:

Study design: The design will be a case-control study. This project is structured around two diagnostic categories: schizophrenia and bipolar I disorder (grouped under the heading "psychotic disorders"). The rationale for grouping schizophrenia and bipolar disorder stems from literature showing a high level of genetic correlation between schizophrenia and bipolar disorder, which may indicate that some of the same genetic variants confer risk for both phenotypes.[13]

Cases will be individuals with a diagnosis of schizophrenia or bipolar disorder, referred to subsequently as psychosis. Controls will be individuals from the same geographic location, without psychosis, who will be matched to cases for age, sex, and ancestry.

Study sites: The study will be conducted over four years, starting in September 2018 and ending in 2022, at several sites in Ethiopia, Kenya, South Africa, and Uganda (Figure 1).[14] DNA

extraction will be performed on site and genotyping will performed at the Broad Institute in the USA using the Illumina Global Screening Array. Sites were selected on the basis of the following criteria: 1) proven track record of psychiatric research; 2) availability of research personnel and the necessary research infrastructure to be able to recruit thousands of participants; 3) existing trusted relationships from prior collaborations. Each of the countries where participants will be recruited from has enormous genetic diversity within and between them, which is likely to improve the ability of this research to answer the study objectives. Pilot studies are currently underway in all four countries.

Ethiopia. Participants will be recruited from the Amanuel Mental Specialized Hospital (cases) and Black Lion Hospital also known as the Tikur Anbessa Hospital (controls).

Kenya. In Kenya, participants will be recruited from the Moi Teaching and Referral Hospital and affiliated sites in Webuye, Kapenguria, Kitale, Kapsabet, Iten, and Kakamega and the KEMRI-Wellcome Trust Research Programme with recruiting sites in Kilifi County, Malindi sub-County, Port Reitz, and Coast General Provincial Hospitals.

South Africa. Participants will be recruited from the Western Cape, Eastern Cape and the Gauteng provinces. In the Western Cape, participants will be recruited from the Valkenberg, Lentegeur, Khayelitsha District, and Groote Schuur Hospitals as well as a number of community clinics. In the Eastern Cape, participants (cases) will be recruited from the Fort England Psychiatric, Elizabeth Donkin, Tower Psychiatric, and Komani Hospitals while controls will be recruited from the Nelson Mandela Academic and Dora Nginza Hospitals as well as five affiliated health clinics.

Uganda. Study participants in Uganda will be recruited from the Butabika National Mental Health Referral Hospital (cases and controls), Naguru (controls only), Arua (cases and controls), Mbarara (cases and controls), and Gulu Regional Referral Hospitals (cases and controls).

Inclusion criteria: Individuals with a clinical diagnosis of psychosis (cases) as confirmed by clinician referral and/or medical record review, and those without a clinical diagnosis of psychosis (controls) will be eligible to participate. All participants (cases and controls) will be required to provide written informed consent or a fingerprint in case of illiteracy and must be at least 18 years old. To ensure that participants (cases and controls) have sufficient capacity and autonomy to consent to the study, we will use the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC),[15] which has been used as an iterative learning tool in similar populations in South Africa.[16]

Exclusion criteria for cases: Individuals (cases) will be excluded if the following are present:

- Absence of a diagnosis of a psychotic disorder
- Severe, intrusive levels of psychiatric symptoms at the time of consent
- Intoxication or withdrawal from alcohol or substance abuse at the time of consent
- A current psychiatric hospitalization (inpatients)
- Involuntary detention at the time of consent
- Lack fluency in one of the languages the consent form has been translated into
- Lack capacity to consent to the study as determined by the UBACC.

Potential cases may be approached at a later date once their symptoms are controlled if they meet other inclusion criteria. There will be no exclusions for cases or controls based on sex, ancestry/ethnicity, religious affiliations, or sexual orientation.

Exclusion criteria for controls: Potential controls will be excluded if they:

- Have current psychotic symptoms or a past diagnosis of a psychotic disorder
- Are currently taking medication for psychosis
- Have acute levels of alcohol or substance misuse as demonstrated by being a current inpatient or under acute medical care for substance misuse
- Are not fluent in any of the languages the consent form has been translated into
- Lack capacity to consent to the study as determined by the UBACC.

Participant recruitment: Eligible participants (cases) will be identified by clinical staff through review of their medical records. A research assistant (RA), who is a nurse, clinical officer, clinical assistant, or bachelor's level accredited RA and who has received study specific training and human subjects training, will approach the prospective participant, carefully explaining to them in the local language the purpose and procedures of the study and emphasizing that participation is entirely voluntary and will not impact any medical care they receive. Languages include: Acholi, Afrikaans, Amharic, English, Kigiryama, Kiswahili, Luganda, Lugbara, Runyankole, and isiXhosa.

After reading the information sheet out loud to the subject, the RA will administer the UBACC, a 10-item questionnaire that evaluates the potential participant's understanding of different components of the study. Each response will be scored on a range of 0-2, with 0 representing no understanding and 2 representing a clear understanding. In the NeuroGAP-Psychosis study, the RA will administer the UBACC over a maximum of four trials. The RA will re-explain and re-administer any items the subject answered incorrectly. The process will end if the full score of 20 is obtained. After the 4th trial, participants who are unable to achieve a score of 14.5, the cut-off originally developed for screening decisional capacity using the UBACC,[15] will be excluded.

The RA will proceed to obtain informed consent from those who express interest in participation, and ask participants to provide saliva into an Oragene tube, from which their DNA will be extracted. To maximize the quality of the collected DNA, participants will be asked not to eat, drink, smoke, or chew gum for 30 minutes prior to providing the sample during the consenting process.

Phenotypic assessments were selected on the basis of the following criteria: 1) the particular domain assessed; 2) cross-cultural validity; 3) non-proprietary measures, when possible; 4) investigators' prior experience using the tool; and 5) the time length of administering the tool. The study will collect the following battery of instruments (see Figure 2):

- Mini International Neuropsychiatric Interview, Standard 7.0.2 for DSM-5 (MINI): modules A, C, K, and O on major depressive episodes, manic and hypomanic episodes, and psychotic disorders and mood disorders with psychotic features, respectively[17]
- Life Events Checklist for DSM-5 (LEC-5): a 17-item scale covering exposures to potentially traumatic events[18]
- Alcohol, Smoking and Substance Involvement Screening Test, version 3.0 (ASSIST): a subset of the ASSIST on substance type and use over a participant's lifetime and over the past three months[19]
- Composite International Diagnostic Interview screener (CIDI): a checklist within the CIDI on chronic physical conditions[20] including diabetes, HIV/AIDS, epilepsy/seizures, and tuberculosis.

Participant phenotypic data will be collected using encrypted tablets, and uploaded to a secure cloud-based server. Other sources of data will include chart reviews to ascertain clinical

diagnoses and medication use (both psychotropic and non-psychotropic) by the participants; as well as measurements of blood pressure, heart rate, height, and weight.

Recruitment of controls will take place from general medical facilities (some of these facilities are located within the premises where cases will be identified). Controls will be approached in the same manner as cases and will likely consist of people who are seeking clinical care for themselves, accompanying a friend or family member to a clinic visit, or picking up a medication refill. In one site, at the KEMRI-Wellcome Trust Research Programme in Kenya, controls may also be recruited from homes through the Kilifi Health and Demographic Surveillance System (KHDSS).[21] [In KHDSS, census data is recorded every four months through household surveys in the catchment area and linked with healthcare data from Kilifi District Hospital.]

The study visit for controls will match the process for cases except for two phenotypic batteries, the K10 and the PSQ (see Figure 3); instead of using the MINI for controls, the RA will administer:

- Kessler Psychological Distress Scale (K10): a 10-item questionnaire to measure anxiety and depression[22]
- Psychosis Screening Questionnaire (PSQ): a brief screening tool for psychotic conditions[23]

All participants will receive compensation to offset the time and inconvenience of participating in the study based on recommendations from the local ethics committees and principal investigators. The estimated length of a study visit is 90 to 120 minutes.

Patient and Public Involvement

Patients' priorities and experiences were taken into account in the design of the study. Prior research and clinical work with people with neuropsychiatric disorders at each of the sites informed the research question and study design, including limiting the length of the study battery, minimizing intrusive sample collection (saliva instead of blood), and incorporating a tool as part of the consent process to assess understanding of the research and study visit. In addition, most of the sites have a precedent of genetics research and we engaged with teams from those other genetic studies to incorporate patient concerns emerging from genetic research. Study findings will be shared with local community advisory boards.

Sample size and power calculation: Over a period of four years, the study team expects to recruit a total of 17,000 schizophrenia and bipolar disorder cases and 17,000 controls across four countries: Ethiopia, Kenya, South Africa, and Uganda. The geographical and demographic sampling distribution of the controls will follow that of the cases, and cases and controls will be matched for ethnic ancestry in all analyses. These numbers, described in more detail below, will give adequate statistical power to independently examine the genetic architecture of schizophrenia and bipolar disorder in African populations, the distribution and influence of rare copy number variation on neuropsychiatric disease risk in African populations, and determine the consistency of genetic influences on neuropsychiatric disease risk between populations of European and African descent. Ultimately these data will be analyzed in combination with other studies of psychiatric genetics, most notably from the Psychiatric Genomics Consortium, which will further increase our power to discover new and validate current variants with small effect sizes.

With 17,000 cases and controls, using a threshold of 2.5×10^{-8} for genome-wide significance, we have >90% power to find common variants (Minor Allele Frequency (MAF) >5%) with an odds ratio of 1.15 or greater.[24] For more common variants with MAF=25%, we have 89% power to a

see significant odds ratio as low as 1.06. These effect sizes are within the range of what was found in a recent GWAS of schizophrenia in a European population.[3]

DNA extraction and processing: All study participants will provide a 2.0 mL sample of saliva by spitting into a funnel in an Oragene(R) OR-500 kit. DNA extraction will take place in-country, with one aliquot of the extract sent to the Broad Institute in Cambridge, MA, USA for genotyping. The remaining DNA will be frozen and stored in-country within designated biobanks and will be governed by the rules of its institution. In Ethiopia, DNA will be stored at the Black Lion Hospital. In Kenya, DNA will be stored at Moi University and KEMRI-Wellcome Trust Research Programme Biorepository Laboratories. In South Africa, DNA will be stored in a biobank at the Division of Human Genetics at the University of Cape Town, while in Uganda, DNA will be stored at the Integrated Biorepository Lab of H3Africa in Uganda (IBR-H3A-U).

All samples will be assayed on the same genotyping array at the same facility. To maximize information and allow for comparisons across studies genotyped on different platforms, we will impute SNP genotypes not present on specific arrays. Genotype imputation of untyped markers will be performed using default parameters in IMPUTE2 using appropriate reference panels.

Training: To ensure research staff understand the study and their responsibilities, teams will undergo multiple trainings. Study staff who have contact with research participants will complete the online Human Subject Training offered by the US National Institutes of Health. Site leadership and Harvard staff will provide in-person training to the teams by providing an overview of the project, leading discussions, providing role playing on ethics and confidentiality, recruiting and consenting participants, on reportable events, and on storing forms and data properly. Additionally, there will be hands-on training using the electronic tablets, collecting saliva using Oragene kits, administering the phenotypic batteries, and taking physiological measures, such as height and weight. A checklist will be created of key skills staff need before they can begin interacting with participants one-on-one (Figure 4). This form will need to be signed and dated by the staff member and the staff member's supervisor before he or she can begin working with participants; it will then be uploaded to the regulatory binder. Harvard teams will reinforce these skills with multiple site visits per year and weekly video calls.

In addition, Broad Institute wet lab staff will work in partnership with the lab teams in Africa to standardize the DNA extraction process across the five sites and to increase the quality of the extractions through visits, video calls, and email. Where needed, the Broad staff will train lab members on DNA extraction from saliva and establish a sample chain of custody process in order to decrease the likelihood of sample swaps.

Statistical analysis plans:

a) Ancestry analyses. In contrast to most samples previously analyzed by the psychiatric genetics community – which were predominately of European ancestry[25] – the NeuroGAP-Psychosis samples will include individuals of predominantly African ancestry and will be drawn from multiple populations across East and South Africa where genetic variation will be more structured and demographically complex.[26 27] Therefore, before proceeding with any phenotype analyses we will first analyze genetic ancestry in our cohorts to identify large, homogeneous groups as described below to conserve optimal power in the GWAS and ultimately control for additional population stratification within these groups to avoid spurious results.

We will take advantage of the large amount of genotypic information available and control for potential population stratification in a standard and validated two-step process. First, we will compute principal components analysis (PCA) to adjust for population substructure. We will also

include principal components (PCs) when computing relatedness using PC-AiR.[28] We will define clusters of ancestries using ADMIXTURE,[29] and choose the optimal number of clusters with 5-fold cross-validation to identify major groups of participants with similar ancestry. The inclusion of extensive population reference samples compiled by our group and available from other studies, such as the 1000 Genomes Project[30] and the African Genome Variation Project,[27] will increase the power of this approach and lead to stable individual ancestry calling across studies. Second, if significant differences are found between populations in our cohorts, as expected, we will define more homogeneous populations within several standard deviations of these reference data, selecting each population by constructing multidimensional ellipses across the PCs. Our initial analyses will then be conducted separately on more homogeneous ancestry groups, and later meta-analyzed. Within each group, we will conduct PCA again using PC-AiR, and the PCs that explain the most variation within the cohort will be included as covariates in the regression analyses. To adjust for cryptic relatedness that considers ancestry, we will compute a genetic relatedness matrix using PC-Relate,[31] which can be used as a random effect in a linear or logistic mixed model. In addition, as new methods are developed by our group and others to account for population structure, we will modify our analyses when appropriate.

b) Genome-wide Association Study (GWAS). The primary GWAS discovery analysis will be conducted for the "psychotic disorders" diagnosis, as well as through subsets stratified by primary disorder (schizophrenia, bipolar disorder) using logistic regression. All analyses will be performed using PLINK[32] and R.[33] To correct for multiple comparisons and the large number of variants in the African genome, we will apply a more conservative significance threshold of $p<2.5x10^{-8}$.[34] Using these subsets, we will assess the genetic correlation and diagnostic stratification among phenotypes.

c) Heritability estimates. Estimation of heritability, the fraction of disease risk attributable to inherited genetic differences, is an important initial analytic goal. Heritability estimates speak to the anticipated architecture and power of a genome-wide association study, a primary endpoint of the NeuroGAP-Psychosis project. We will use Genome-wide Complex Trait Analysis (GCTA) to estimate SNP-based heritability within Africa.[25 35] We will assess genetic correlation with existing GWAS in other globally diverse populations using POPCORN.[36]

ETHICS AND DISSEMINATION:

Ethics: Ethical and safety considerations will be taken across multiple levels. Since the subjects the study aims to recruit are deemed vulnerable populations, additional measures will be taken to protect them. Potential participants will be excluded if they are presenting with severe, intrusive levels of psychiatric symptoms at the time of consent. In addition, as described previously, the RAs will use the UBACC [15 16] during the consent process to make sure participants understand the study, what is required of them, and that they can withdraw at any point. Participants who pass the UBACC and who want to continue will be required to provide written informed consent or a fingerprint in lieu of a signature.

All steps will be taken to mitigate the potential loss of confidentiality for participants. A two-part ID system will be instituted in order to limit the possibility of connecting the participant with their data. There will be one ID for the phenotypic information, and one for genetic information. A single encrypted database linking the phenotypic ID and genetic ID will be stored in-country. The participant's name and contact information will be on the consent form, which will be stored in a locked cabinet. This information will be put in a separate secure database and will never leave the country of origin. No data that is considered identifiable, by Health Insurance Portability and Accountability Act (HIPPA) standards, will ever be collected on the electronic tablets.

Ethical clearances to conduct this study have been obtained from all participating sites, including:

- Ethiopia: Addis Ababa University College of Health Sciences (#014/17/Psy) and the Ministry of Science and Technology National Research Ethics Review Committee (#3.10/14/2018).
- Kenya: Moi University School of Medicine Institutional Research and Ethics Committee (#IREC/2016/145, approval number: IREC 1727), Kenya National Council of Science and Technology (#NACOSTI/P/17/56302/19576) KEMRI Centre Scientific Committee (CSC# KEMRI/CGMRC/CSC/070/2016), KEMRI Scientific and Ethics Review Unit (SERU# KEMRI/SERU/CGMR-C/070/3575)
- South Africa: The University of Cape Town Human Research Ethics Committee (#466/2016) and Walter Sisulu University Research and Ethics Committee (# 051/2016)
- Uganda: The Makerere University School of Medicine Research and Ethics Committee (SOMREC #REC REF 2016-057) and the Uganda National Council for Science and Technology (UNCST #HS14ES)
- USA: The Harvard T.H. Chan School of Public Health (#IRB17-0822)

Dissemination: We aim to share results through a number of mechanisms in addition to conferences and peer-reviewed publications. In order to maximize scientific utility of the samples/data and to minimize data waste, both genomic and phenotypic data from NeuroGAP-Psychosis will be deposited in controlled access public databases, such as the database of Genotypes and Phenotypes (dbGaP), the Psychiatric Genomics Consortium (PGC), and/or the European Genome-Phenome Archive (EGA). By pooling NeuroGAP-Psychosis' data with that from other studies, it will be possible to maximize statistical power. Sharing the data will also contribute to the "genomic revolution" currently underway in Africa due to the pioneering work of the African Genome Variation Project, MalariaGEN, African Society of Human Genetics, and Human Hereditary and Health in Africa (H3Africa).[27 37-40]

To help ensure all collaborators have the capacity to work with all the data that will be created, a unique training program has been created, the Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER),[41] that will run in parallel to the NeuroGAP-Psychosis study. Over a two-year period, 17 early-career researchers from Ethiopia, Kenya, South Africa, and Uganda will take part in online and in-person trainings on topics including biostatistics, genetic analysis, and epidemiology. Ultimately, the goal of GINGER is to build the next generation of neuropsychiatric geneticists who will be able to work on the NeuroGAP-Psychosis data and continue its legacy in the future.

Authors' contributions:
KK and DS conceptualized and designed the study.
AS, DA, RS, LA, EK, SK, VdM, CN, DS, ST, ZZ, and KK wrote the protocol, selected the phenotype questionnaires, and added site-specific information.
LC and ARM wrote the sample size, power calculations, and statistical analysis plans.
MC and GS contributed to the UBACC process.
AS and DA wrote the draft of the manuscript and incorporated the revisions by the co-authors; they contributed equally to this paper.
All authors reviewed the manuscript for intellectual content, contributed to revisions, and approved the final version for publication.

Funding statement: This study is funded by the Stanley Center for Psychiatric Research at Broad Institute. The sponsor was involved in the study design, in the writing of the report, and in the decision to submit the paper for publication. The sponsor will be involved in the analysis and interpretation of the data. A.R.M. was supported by funding from the National Institutes of Health (K99MH117229).

Competing interests statement. Dan Stein has received research grants and/or consultancy honoraria from Biocodex, Lundbeck, Servier, and Sun.

Legends for Figures:

Figure 1: Proposed collection sites for NeuroGAP-Psychosis

Figure 2: Phenotyping tools for study participants

Figure 3: The study process for cases and controls

Figure 4: NeuroGAP-Psychosis Training Checklist

REFERENCES:

1. Whiteford HA, Ferrari AJ, Degenhardt L, et al. Global Burden of Mental, Neurological, and Substance Use Disorders: An Analysis from the Global Burden of Disease Study 2010. In: Patel V, Chisholm D, Dua T, et al., eds. Mental, Neurological, and Substance Use Disorders: Disease Control Priorities, Third Edition (Volume 4). Washington (DC): The International Bank for Reconstruction and Development / The World Bank
(c) 2016 International Bank for Reconstruction and Development / The World Bank. 2016.
2. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet (London, England)* 2013;382(9904):1575-86. doi: 10.1016/s0140-6736(13)61611-6 [published Online First: 2013/09/03]
3. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511:421. doi: 10.1038/nature13595 https://www.nature.com/articles/nature13595#supplementary-information
4. Pardinas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics* 2018;50(3):381-89. doi: 10.1038/s41588-018-0059-2 [published Online First: 2018/02/28]
5. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. *Pharmacogenomics* 2009;10(2):191-201. doi: 10.2217/14622416.10.2.191 [published Online First: 2009/02/12]
6. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016;538(7624):161-64. doi: 10.1038/538161a [published Online First: 2016/10/14]
7. Li Z, Chen J, Yu H, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature genetics* 2017;49(11):1576-83. doi: 10.1038/ng.3973 [published Online First: 2017/10/11]
8. Dalvie S, Koen N, Duncan L, et al. Large Scale Genetic Research on Neuropsychiatric Disorders in African Populations is Needed. *EBioMedicine* 2015;2(10):1259-61. doi: 10.1016/j.ebiom.2015.10.002 [published Online First: 2015/12/03]
9. Martin AR, Gignoux CR, Walters RK, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American journal of human genetics* 2017;100(4):635-49. doi: 10.1016/j.ajhg.2017.03.004 [published Online First: 2017/04/04]
10. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015;526(7571):68-74. doi: 10.1038/nature15393 [published Online First: 2015/10/04]
11. Wojcik G, Graff M, Nishimura KK, et al. Genetic Diversity Turns a New PAGE in Our Understanding of Complex Traits. *bioRxiv* 2017 doi: 10.1101/188094
12. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual review of genomics and human genetics* 2008;9:403-33. doi: 10.1146/annurev.genom.9.081307.164258 [published Online First: 2008/07/03]
13. Cardno AG, Owen MJ. Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. *Schizophrenia bulletin* 2014;40(3):504-15. doi: 10.1093/schbul/sbu016 [published Online First: 2014/02/26]
14. Ulrich T. Nations Represented in NeuroGAP-Psychosis: Broad Communications, 2018.
15. Jeste DV, Palmer BW, Appelbaum PS, et al. A new brief instrument for assessing decisional capacity for clinical research. *Arch Gen Psychiatry* 2007;64(8):966-74. doi: 10.1001/archpsyc.64.8.966 [published Online First: 2007/08/08]
16. Campbell MM, Susser E, Mall S, et al. Using iterative learning to improve understanding during the informed consent process in a South African psychiatric genomics study. *PLOS ONE* 2017;12(11):e0188466. doi: 10.1371/journal.pone.0188466
17. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of clinical psychiatry* 1998;59 Suppl 20:22-33. [published Online First: 1999/01/09]
18. Weathers FW, Blake DD, Schnurr PP, et al. The Life Events Checklist for DSM-5 (LEC-5). *Instrument available from the National Center for PTSD wwwptsdvagov* 2013

19. Humeniuk R, Henry-Edwards S, Ali R, et al. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): manual for use in primary care. 2010
20. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International journal of methods in psychiatric research* 2004;13(2):93-121. [published Online First: 2004/08/07]
21. Scott JA, Bauni E, Moisi JC, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *International journal of epidemiology* 2012;41(3):650-7. doi: 10.1093/ije/dys062 [published Online First: 2012/05/01]
22. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. *Archives of general psychiatry* 2003;60(2):184-9. [published Online First: 2003/02/13]
23. Bebbington P, Nayani T. Psychosis Screening Questionnaire. *International journal of methods in psychiatric research* 1995;5(1): 11-19
24. Skol AD, Scott LJ, Abecasis GR, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature genetics* 2006;38(2):209-13. doi: 10.1038/ng1706 [published Online First: 2006/01/18]
25. de Candia TR, Lee SH, Yang J, et al. Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. *American journal of human genetics* 2013;93(3):463-70. doi: 10.1016/j.ajhg.2013.07.007 [published Online First: 2013/08/21]
26. Busby GBJ, Band G, Si Le Q, et al. Admixture into and within sub-Saharan Africa. *eLife* 2016;5:e15266. doi: 10.7554/eLife.15266
27. Gurdasani D, Carstensen T, Tekola-Ayele F, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* 2015;517(7534):327-32. doi: 10.1038/nature13997 [published Online First: 2014/12/04]
28. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genetic epidemiology* 2015;39(4):276-93. doi: 10.1002/gepi.21896 [published Online First: 2015/03/27]
29. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 2009;19(9):1655-64. doi: 10.1101/gr.094052.109 [published Online First: 2009/08/04]
30. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491(7422):56-65. doi: 10.1038/nature11632 [published Online First: 2012/11/07]
31. Conomos MP, Reiner AP, Weir BS, et al. Model-free Estimation of Recent Genetic Relatedness. *American journal of human genetics* 2016;98(1):127-48. doi: 10.1016/j.ajhg.2015.11.022 [published Online First: 2016/01/11]
32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007;81(3):559-75. doi: 10.1086/519795 [published Online First: 2007/08/19]
33. R: A Language and Environment for Statistical Computing [program]. Vienna, Austria, 2018.
34. Kanai M, Tanaka T, Okada Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *Journal of human genetics* 2016;61(10):861-66. doi: 10.1038/jhg.2016.72 [published Online First: 2016/10/26]
35. Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* 2011;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011 [published Online First: 2010/12/21]
36. Brown BC, Ye CJ, Price AL, et al. Transethnic Genetic-Correlation Estimates from Summary Statistics. *American journal of human genetics* 2016;99(1):76-88. doi: 10.1016/j.ajhg.2016.05.001 [published Online First: 2016/06/21]
37. Beiswanger CM, Abimiku A, Carstens N, et al. Accessing Biospecimens from the H3Africa Consortium: Biopreserv Biobank. 2017 Apr 1;15(2):95-8. Epub 2017 Apr 1 doi:10.1089/bio.2017.0008.
38. Consortium THA. Enabling the genomic revolution in Africa: H3Africa is developing capacity for health-related genomics research in Africa. *Science (New York, NY)* 2014;344(6190):1346-8. doi: 10.1126/science.1251546
39. Mulder N, Abimiku Al, Adebamowo SN, et al. H3Africa: current perspectives. *Pharmacogenomics and Personalized Medicine* 2018;11:59-66. doi: 10.2147/PGPM.S141546

40. Band G, Rockett KA, Spencer CC, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* 2015;526(7572):253-7. doi: 10.1038/nature15390 [published Online First: 2015/09/30]
41. van der Merwe C, Mwesiga EK, McGregor NW, et al. Advancing neuropsychiatric genetics training and collaboration in Africa. *The Lancet Global health* 2018;6(3):e246-e47. doi: 10.1016/s2214-109x(18)30042-1 [published Online First: 2018/02/13]

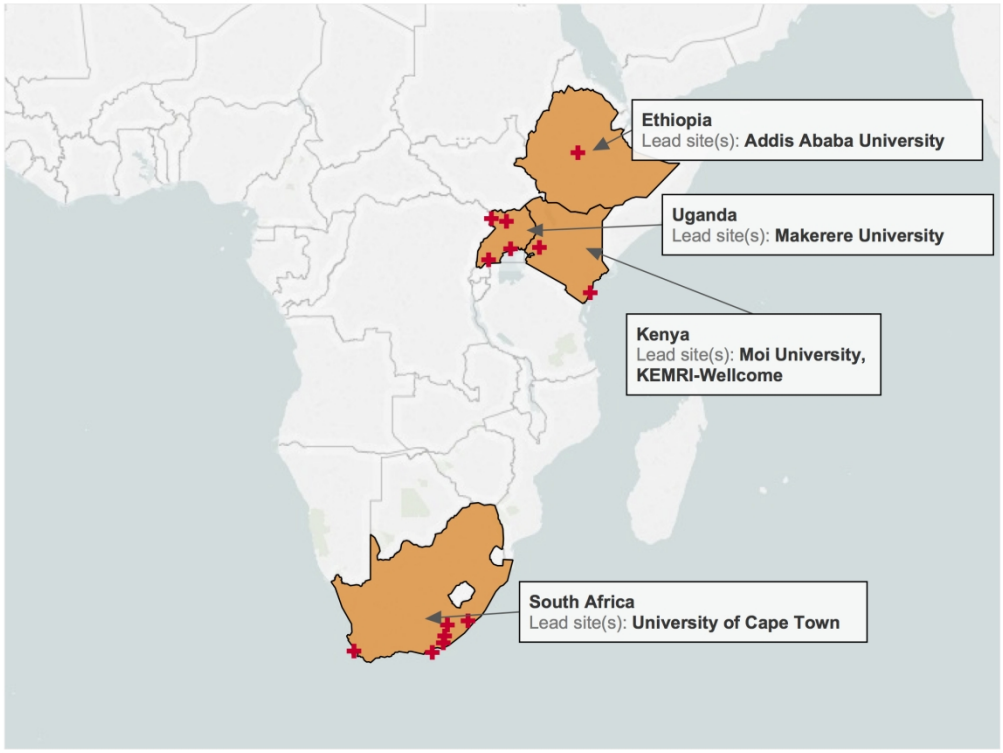

Figure 1: Proposed collection sites for NeuroGAP-Psychosis

192x145mm (300 x 300 DPI)

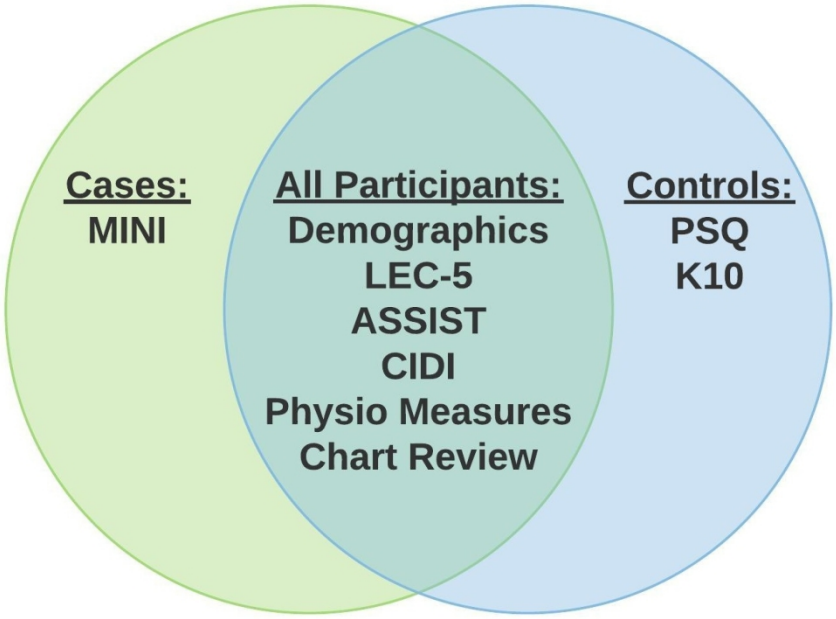

Figure 2: Phenotyping tools for study participants

130x97mm (300 x 300 DPI)

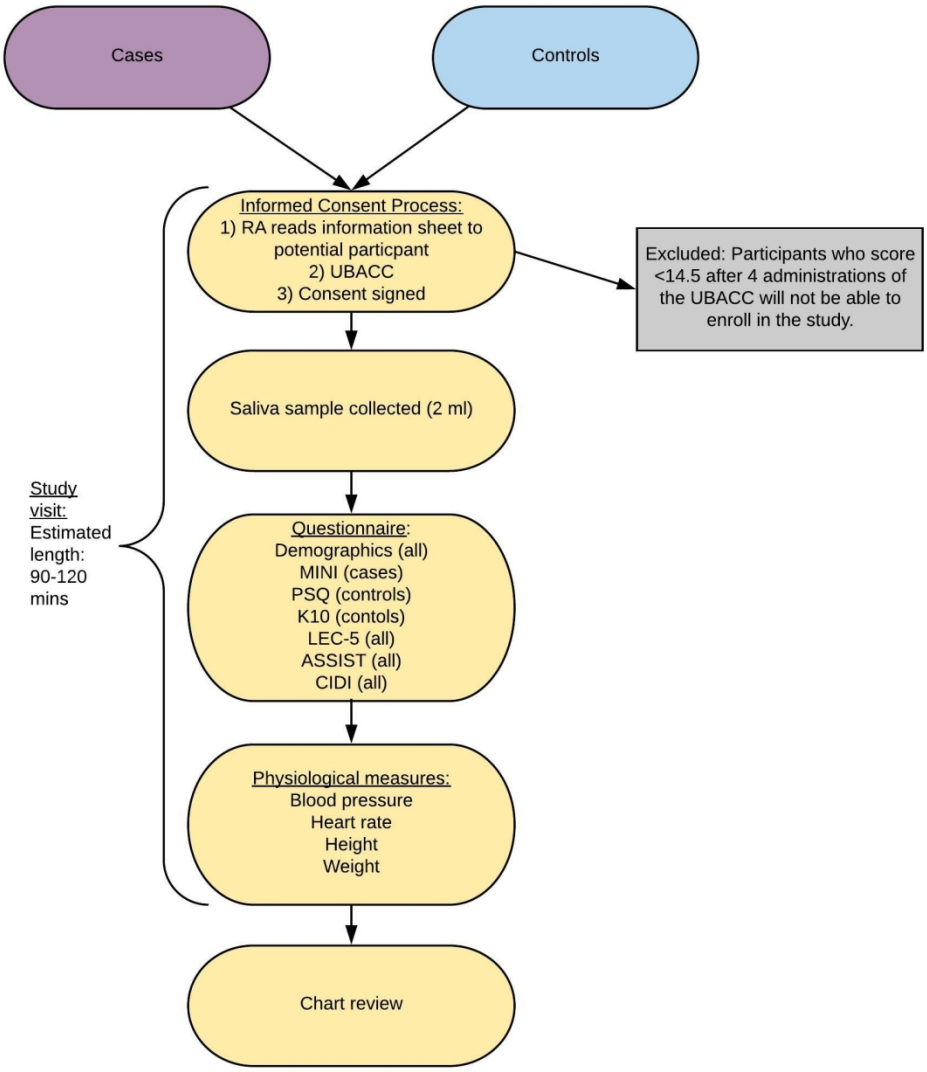

Figure 3: The study process for cases and controls

155x177mm (300 x 300 DPI)

NeuroGAP-Psychosis Training Checklist
For All Study Staff

_____
Print name above

Instructions: Check the boxes when you have completed each task. When you are done, you and your supervisor must sign and date the form to attest that you have completed the training and have approval to start working with NeuroGAP-Psychosis subjects. If some of these components are not applicable, the PI should write "Not Applicable" next to the item.

☐ Online US National Institutes of Health Ethics Training Date completed:

☐ CV on file - must be signed and dated

☐ Trained on the NeuroGAP-Psychosis Protocol

Consent Process

☐ Study Information Sheet

☐ Consent Form

☐ UBACC

Study Tools

☐ MINI ☐ LEC-5

☐ K10 ☐ PSQ

☐ ASSIST ☐ CIDI

Physiological Measures

☐ Blood pressure, heart rate, weight, height

Role Playing

☐ Role played all of the study activities from beginning to end with other members of the NeuroGAP-Psychosis team

Shadowing

☐ Watched an experienced member of the study team complete 5 study visits with participants

ATTESTATION:

_____ _____
Signature of study team member Signature of supervisor or PI

_____ _____
Print name of study team member Print name of supervisor or PI

_____ _____
Date Date

Figure 4: NeuroGAP-Psychosis Training Checklist

196x203mm (300 x 300 DPI)