PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa, and Uganda
AUTHORS	Stevenson, Anne; Akena, Dickens; Stroud, Rocky; Atwoli, Lukoye; Campbell, Megan; Chibnik, Lori; Kwobah, Edith; Kariuki, Symon; Martin, Alicia; de Menil, Victoria; Newton, Charles; Sibeko, Goodman; Stein, Dan; Teferra, Solomon; Zingela, Zukiswa; Koenen, Karestan

VERSION 1 – REVIEW

REVIEWER	Daniel Smith University of Glasgow, UK.
REVIEW RETURNED	30-Jul-2018

GENERAL COMMENTS	NeuroGAP-Psychosis is an excellent initiative and this protocol paper clearly describes how the study will be conducted. The need for this important study, the objectives and the strengths of the study design are all clearly articulated by the authors. The ethics permissions are all in place and the plan for data analysis and future dissemination is excellent. I have only a few minor queries for clarification: 1. All of the proposed phenotypic assessments of cases are appropriate but I wondered if there should also be a record of current and previous medication use (both psychotropics and non-psychotropics)? 2. What will the thresholds scores be on the K10 and PSQ for excluding potential controls? 3. All participants will receive compensation - is it possible to give an indication of how much this is likely to be (I appreciate it may vary across countries)? 4. Could the authors provide more detail on how controls will be approached and recruited, beyond coming from the same geographic population? Also, controls will be matched for age, sex and ancestry but what about matching also for socioeconomic status and educational level/attainment?

REVIEWER	David Howard University of Edinburgh, UK
REVIEW RETURNED	15-Aug-2018

GENERAL COMMENTS	General Comment The described study protocol is to conduct a genome-wide association study of psychosis using 34,000 individuals and a case-control study design in African populations. The ethical and practical considerations of conducting such a study are well

	considered and the ascertainment of a psychosis phenotype appears robust. This research has many potential benefits, not only for the African populations studied but also may reveal variance components that are relevant to the global population. The reporting of the collection of this data is highly valuable and would be strengthened by adding the collection of additional health and baseline measures. Comment #1 In the participant recruitment section, it states that blood pressure, heart rate, height, and weight will be collected. Have the author's considered in extending this to include other information including other baseline data as well as other diseases and conditions? Would it be practical/possible to obtain access to the medical records of participants? Comment #2 Is it possible to also ascertain whether participants with psychosis are classified as having either bipolar disorder or schizophrenia? There is undoubtedly a shared genetic component underlying both of these disorders, but it would also be potentially valuable to also assess the differences between them and then the disorders individually. During the phenotypic assessments for cases would it be possible to obtain an age of onset for the disorder? Comment #3 It would be useful to provide an indication of the likely reference panels which will be used for imputation. It is a little out of date but the Howie, Marchini and Stephens paper may be of use: https://dx.doi.org/10.1534%2Fg3.111.001198 Comment #4 Further information should be provided on how the authors intend to bring together groups with potentially very different ancestries Minor comment #1 Within Exclusion criteria for cases: remove "of" from "withdrawal from of alcohol or substance abuse" Minor comment #2 Within Sample size and power calculation: change "Psychiatric Genetics Consortium" to "Psychiatric Genomics Consortium" Minor comment #3 Within Sample size and power calculation: For the effect sizes (1.15 and 1.06) are these odds ratios or genotype risk ratios? Minor comment #4 Within Ancestry analysis: It is stated 'The inclusion of extensive population reference samples compiled by our group and available from other studies'. Please provide references for these other studies. Minor comment #5 In a couple of places, the p-value threshold for significance is given as < 2.5 × 10-8, the typically applied threshold in genome-wide association studies is < 5 × 10-8. Minor comment #6

	Within the Heritability estimates section: Please state that it is the "SNP-based heritability" you are planning to calculate.

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 1
Reviewer Name: Daniel Smith
Institution and Country: University of Glasgow, UK.
Please state any competing interests or state 'None declared': None declared

NeuroGAP-Psychosis is an excellent initiative and this protocol paper clearly describes how the study will be conducted. The need for this important study, the objectives and the strengths of the study design are all clearly articulated by the authors. The ethics permissions are all in place and the plan for data analysis and future dissemination is excellent. I have only a few minor queries for clarification:
1. All of the proposed phenotypic assessments of cases are appropriate but I wondered if there should also be a record of current and previous medication use (both psychotropics and non-psychotropics)?

We collect current medications for both psychotropics and non-psychotropics (see list below).
Current Medications
• Is the subject currently taking any psychiatric medications? Yes / No
• How many psychiatric medications is the patient currently on? 1 2 3 (circle)

Medication Name

• alprazolam, artest (sedative)
• midazolam, dormicum (sedative)
• zoplicone, Imovane (sedative)
• flunitrazepam, rohypnols (sedative)
• lorazepam (sedative)
• zolpidem, Stillnox/Ambien (sedative/hypnotic)
• methylphenidate, ritalin (ADHD)
• atomoxetine, Strattera (ADHD)
• donepezil, arisept (Alzheimers)
• phenobarbital (anticonvulsant)
• sodium valproate, Encorate/Epilim (anticonvulsant)
• phenytoin, Epanutin (anticonvulsant)
• gabapentin, neurontin (anticonvulsant)
• lamotrigine, lamictal (anticonvulsant)
• carbamazepine, Tegretol (anticonvulsant)
• Citalopram, Citopam (antidepressant)
• mirtazapine (antidepressant),
• clonazepam, Nexito (antidepressant)
• fluoxetine, Prodep/Prozac/Salipax (antidepressant)
• Venlafaxine, Ventab (antidepressant)
• sertraline, Zosert (antidepressant)
• benzhexol (antiparkinsonian),
• procyclidine, kemadrine (antiparkinsonian)

• chlorpromazine (antipsychotic),
• zuclopenthixol, Clopixol/acuphase (antipsychotic)
• flupenthixol, Fluanoxol (antipsychotic)
• fluphenazine decanoate, modecate (antipsychotic)
• clozapine, Leponex (antipsychotic)
• olanzapine, Olencip/Ozitas (antipsychotic)
• quetiapine, seroquel (antipsychotic)
• risperidone, risdone/Sizodon (antipsychotic)
• haloperidol, Senorm (antipsychotic)
• Amisulpride, Solian (antipsychotic)
• olanzapine, Zyprexa (antipsychotic)
• trifluoperazine, Stelazine (antipsychotic)
• diazepam, Valium (anxiolytic)
• lithium, camcolit (mood stabilizer)
• Other
o If other, Medication Name _____

Non-Psychiatric Medication Categories (Click as many options as apply):

 Diabetes
 Asthma
 Rheumatism
 HIV
 TB
 Hypertension
 Cholesterol
 Contraception
 Auto-immune Disorder
 Epilepsy
 Other (please specify)
 Non-Psychiatric: Other Medications _____
o If other, what type of non-psychiatric medication category? _____

It is not feasible to consistently collect previous medication use in the participant population. There are no electronic medical records in the clinics from which we recruit and the patient files that the study teams have access to are only the most recent ones.

2. What will the thresholds scores be on the K10 and PSQ for excluding potential controls?
There is no exclusion during recruitment or study visit for potential controls for K10 or PSQ scores. We monitor these scores to ensure we are not recruiting controls with severe psychopathology or psychosis – and our pilot data suggests we are not. In our preliminary data, 88.3% of controls scored below a 20 on the Kessler Psychological Distress Scale (K10), which corresponds to overall good mental health and low rates of mental disorder. Similarly, 91% of controls have screened negative for psychotic disorders on the Psychosis Screening Questionnaire (PSQ). We will have the option of excluding some controls during the analyses phase if needed. The determination of thresholds will be made in consultations with local experts as well as consider those in the literature.

3. All participants will receive compensation - is it possible to give an indication of how much this is likely to be (I appreciate it may vary across countries)?
South African participants will be paid 125 Rand (USD $8). Study participants in Western Kenya will be reimbursed for transportation costs only, following the IRB guidelines at Moi University (everyone

receives 1000 Kenyan Shillings). Participants in Eastern Kenya recruited through the KEMRI-Wellcome Trust Research Program will receive 900 Kenyan Shillings (USD $8.5). Ugandan study participants will receive 10,000 Ugandan Shillings (USD $3). In Ethiopia, participants will be paid 100 Ethiopian Birr (USD $4.50).

4. Could the authors provide more detail on how controls will be approached and recruited, beyond coming from the same geographic population?
We will ascertain controls from persons who present for treatment of general medical conditions at university-affiliated general medical hospitals and clinics that draw from similar catchment areas to the psychiatric facilities. The potential controls will likely be people who are seeking clinical care for themselves, accompanying a friend or family member to a clinic visit, or picking up a medication refill. Flyers containing brief information about the study, inclusion/exclusion criteria, and study contact information will be handed out by the research staff to potential participants waiting in the waiting area. The local study team will be the first point of contact for the controls. The study team will approach the potential participant as he/she queues for care or waits in the waiting area, introduce themselves, explain the study and find out if the person would like to learn more. If the patient is interested, he/she will be brought to a private room where the study will be explained in more detail to determine if the person would like to participate.

Additional text has been added to the manuscript to describe the recruitment of controls.

Also, controls will be matched for age, sex and ancestry but what about matching also for socioeconomic status and educational level/attainment?
We appreciate the reviewer bringing up this point. While we will not directly match on SES or education, cases and controls are ascertained from the same ascertainment areas, clinic populations, and language groups which helps assure the cases and controls will be similar on those factors. We will be collecting information on years of schooling in order to consider this if necessary in the analysis.

Reviewer: 2
Reviewer Name: David Howard
Institution and Country: University of Edinburgh, UK
Please state any competing interests or state ‘None declared’: None

General Comment
The described study protocol is to conduct a genome-wide association study of psychosis using 34,000 individuals and a case-control study design in African populations. The ethical and practical considerations of conducting such a study are well considered and the ascertainment of a psychosis phenotype appears robust. This research has many potential benefits, not only for the African populations studied but also may reveal variance components that are relevant to the global population. The reporting of the collection of this data is highly valuable and would be strengthened by adding the collection of additional health and baseline measures.

Comment #1
In the participant recruitment section, it states that blood pressure, heart rate, height, and weight will be collected. Have the author’s considered in extending this to include other information including other baseline data as well as other diseases and conditions? Would it be practical/possible to obtain access to the medical records of participants?

Yes, we have access to medical records for cases and for some controls. Please refer to Reviewer #1, Question #1 for more detail. We collect baseline data through both the chronic disease section of the CIDI and through medical records on medications taken (psychiatric and non-psychiatric) and

conditions including diabetes, HIV, epilepsy, cerebral malaria, neurosyphilis, brain trauma etc. Additional clarification has been added to the manuscript on access to the data we will be able to collect through the chronic disease section of the CIDI and from chart review.

Comment #2
Is it possible to also ascertain whether participants with psychosis are classified as having either bipolar disorder or schizophrenia? There is undoubtedly a shared genetic component underlying both of these disorders, but it would also be potentially valuable to also assess the differences between them and then the disorders individually.

Our Research Assistants (RAs) review the medical chart for every case that we enroll into our study. We capture individual diagnoses, including schizophrenia and bipolar disorder, and will be able to analyze the data at the individual diagnosis level or group "psychosis" level.
As the reviewer is aware, the landscape of psychiatric genetics is constantly evolving. We are aiming to collect our data in a manner that will allow flexibility in the analysis and use state-of-the-art methods at the time we conduct the analysis. Thus, we can do either/both. That is, run GWAS on the samples as a whole (all psychosis disorders) as well as subsets stratified by primary disorder (schizophrenia, bipolar disorder, etc). Using these subsets, we will assess the genetic correlation and diagnostic stratification among phenotypes.

During the phenotypic assessments for cases would it be possible to obtain an age of onset for the disorder?

Our initial assessment did try and obtain the onset of the psychosis disorder. However, the collection of this information proved to be difficult, unreliable, and inconsistent so we discontinued trying to obtain this information. A few reasons for this included: lack of electronic medical record keeping systems and lack of clarity for participants as to what onset means. For example, participants endorsed "onset" as first time being admitted to hospital and not the first episode when the symptoms started, and participants having had visited multiple clinics and endorsing "onset" as first time visiting the clinic they are being recruited from.

Comment #3
It would be useful to provide an indication of the likely reference panels which will be used for imputation. It is a little out of date but the Howie, Marchini and Stephens paper may be of use: https://dx.doi.org/10.1534%2Fg3.111.001198.
We will be using the 1000 Genomes Project (Abecasis et al. 2012) and African Genomic Variation Project (Gurdasani et al. 2014) as reference panels. This has been added to the manuscript.

Comment #4
Further information should be provided on how the authors intend to bring together groups with potentially very different ancestries
We agree that there will be greater heterogeneity in our study than previous Eurocentric studies, which will aid the identification of novel associations and fine-mapping. To ensure this increased heterogeneity does not simultaneously increase false discovery rates, we will assess population structure using principal components analysis (PCA) and ADMIXTURE of the NeuroGAP genetic data with additional reference data from the African Genome Variation Project and 1000 Genomes Project. Guided by these reference panels and as in previous studies of large biobank data with globally diverse ancestries (e.g. the UK Biobank, as referred to here: https://github.com/Nealelab/UK_Biobank_GWAS), we will define homogeneous ancestries within several standard deviations of these reference data, selected using multidimensional ellipses across the first 20 PCs. Within each of the ancestry subsets, we will rerun PCA and compute a genetic relationship matrix that accounts for these PCs to adjust for cryptic relatedness as a random effect in

a linear or logistic mixed model. We have added this detail to the manuscript under ancestry analyses.

Minor comment #1
Within Exclusion criteria for cases: remove “of” from “withdrawal from of alcohol or substance abuse”
Thank you for pointing this out. This has been corrected.

Minor comment #2
Within Sample size and power calculation: change “Psychiatric Genetics Consortium” to “Psychiatric Genomics Consortium”
Thank you for catching this. This has been corrected.

Minor comment #3
Within Sample size and power calculation: For the effect sizes (1.15 and 1.06) are these odds ratios or genotype risk ratios?
We thank the reviewer for pointing out this omission. The effect sizes used for the power calculations are odds ratios. This is now specified in the paper.
“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 an odds ratio of 1.15 or greater. For more common variants with MAF=25%, we have 89% power to a see significant effects 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.”

Minor comment #4
Within Ancestry analysis: It is stated ‘The inclusion of extensive population reference samples compiled by our group and available from other studies’. Please provide references for these other studies.
References to the African Genome Variation Project and the 1000 Genome Project have been added to the text.

Minor comment #5
In a couple of places, the p-value threshold for significance is given as < 2.5 × 10-8, the typically applied threshold in genome-wide association studies is < 5 × 10-8.
We agree with the reviewer that this should be clarified. To control for the increased number of variants in the African genome we typically use a more conservative threshold of 2.5 x 10-8 (Kanai, 2016). Higher rates of genetic diversity also result in a larger number of effective tests, meaning that the standard multiple testing threshold of p < 5e-8 needs to be roughly twice as stringent in African GWAS (p < 2.5e-8). We have clarified this in the text.

“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 PLINK32 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.”

Minor comment #6
Within the Heritability estimates section: Please state that it is the "SNP-based heritability" you are planning to calculate.
This has been clarified in the text.
References

Abecasis, G. R., A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, and G. A. McVean. 2012. 'An integrated map of genetic variation from 1,092 human genomes', Nature, 491: 56-65.
Gurdasani, Deepti, Tommy Carstensen, Fasil Tekola-Ayele, Luca Pagani, Ioanna Tachmazidou, Konstantinos Hatzikotoulas, Savita Karthikeyan, Louise Iles, Martin O. Pollard, Ananyo Choudhury, Graham R. S. Ritchie, Yali Xue, Jennifer Asimit, Rebecca N. Nsubuga, Elizabeth H. Young, Cristina Pomilla, Katja Kivinen, Kirk Rockett, Anatoli Kamali, Ayo P. Doumatey, Gershim Asiki, Janet Seeley, Fatoumatta Sisay-Joof, Muminatou Jallow, Stephen Tollman, Ephrem Mekonnen, Rosemary Ekong, Tamiru Oljira, Neil Bradman, Kalifa Bojang, Michele Ramsay, Adebowale Adeyemo, Endashaw Bekele, Ayesha Motala, Shane A. Norris, Fraser Pirie, Pontiano Kaleebu, Dominic Kwiatkowski, Chris Tyler-Smith, Charles Rotimi, Eleftheria Zeggini, and Manjinder S. Sandhu. 2014. 'The African Genome Variation Project shapes medical genetics in Africa', Nature, 517: 327.

VERSION 2 – REVIEW

REVIEWER	David Howard University of Edinburgh, UK.
REVIEW RETURNED	26-Nov-2018

GENERAL COMMENTS	I am satisfied with the authors' replies to my comments and I am supportive of publication.