Genetic Determinants Of Major Depressive Disorder (MDD) Among Adult Persons Living With HIV In Uganda.

Olga Nsangi Tendo[1,2*], Ronald Galiwango[1,2,8], Eugene Kinyanda[3], Martha Sajatovic[7], Mark Kaddumukasa[4], Martin Kaddumukasa[4], Elly Katabira[4], Catherine Nabbumba[6], Seedat Soraya[5,6], Sian Hemmings[5,6] and Allan Kalungi[3]

1. Department of Immunology and Molecular Biology, Makerere University, P.O.Box 7072, Kampala, Uganda
2. African Centers of Excellence in Bioinformatics and Data Intensive Sciences, Kampala, Uganda
3. Medical Research Council/Uganda Virus Research Institute (MRC/UVRI) and London School of Hygiene & Tropical Medicine (LSHTM) Uganda Research Unit, Entebbe, Uganda
4. Department of Medicine, Makerere University, P.O.Box 7072, Kampala, Uganda
5. South African Medical Research Council/Stellenbosch University Genomics of Brain Disorders Unit, Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
6. Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
7. School of Medicine, Case Western Reserve University School, Cleveland, Ohio, United States of America
8. Infectious Diseases Institute, College of Health Sciences, Makerere University, P.O Box 22418, Kampala, Uganda

* Correspondence: tendoolga@gmail.com; Tel.: +256 7055033103

Abstract

Background

Major Depressive Disorder (MDD) has a heritable component, with estimates of heritability ranging from 30% to 40%. Depression is a significant comorbidity in people living with HIV (PLWHIV), increasing the risk of suicide-related behaviors. This study investigated the genetic risk loci associated with MDD among adults living with HIV in Uganda, where limited data exist on this relationship.

Methods

The case-control study analyzed 282 samples (139 MDD cases and 143 controls), assessed for MDD at baseline, six months, and one year using the Mini International Neuropsychiatric Interview. Blood samples were collected at these intervals, with DNA genotyping conducted in South Africa using the H3Africa array. Data were analyzed using PLINK2 and GEMMA for quality control and genome-wide association analysis respectively, followed by functional mapping with FUMA.

Results

While no significant single nucleotide polymorphisms (SNPs) were identified at the genome-wide threshold, six SNPs were found to be suggestively associated with MDD. These SNPs, which have been associated with other psychiatric conditions like Alzheimer's, alcohol use disorder, and bipolar disorder and have not previously been linked to MDD.

Conclusion

The study suggests the potential for novel MDD genetic risk loci discovery in PLWHIV and people of African ancestry, especially with larger sample sizes.

Keywords: major depressive disorder, single nucleotide polymorphism, risk loci, genome-wide association study.

1. Introduction

Major depressive disorder (MDD) is a common, disabling neuropsychiatric disorder, with multifactorial etiology [1]. The disorder is responsible for 50 million years lived with disability (YLD) worldwide, accounting for 7.5% of global total YLD, and is therefore regarded as the largest contributor to non-fatal health loss [2]. Family and twin studies indicate a strong genetic contribution to the etiology of MDD, with heritability estimates of 30% to 40% reported [3].

MDD is a common cause of psychiatric morbidity in persons living with human immunodeficiency virus (PLWHIV) [4, 5], with 39% of HIV patients having MDD [6]. In 2019, there were 25.6 million PLWHIV in sub-Saharan Africa (SSA), with the rate of depression being between 9 and 32% [7]. MDD has been associated with negative behavioral and clinical outcomes in HIV-affected individuals, including rapid HIV disease progression and mortality, poor adherence to antiretroviral therapy, risky sexual behavior and decreased utilization of health facilities [8, 9] . Despite the impact of MDD on PLWHIV, it has been largely overlooked as a priority within integrated HIV care services in sub-Saharan Africa [5]. Previous studies have shown that depression is not only associated with higher HIV viral loads and lower CD4 cell counts [10] but also hastens the progression to AIDS and elevates the risk of mortality [10]. Furthermore, depression has been reported to reduce adherence to antiretroviral therapy (ART), weaken its therapeutic effects, and compromise the medication outcomes at both individual and population levels. Adherence to ART medications is instrumental in treatment effectiveness and clinical outcomes, while ART interruption and discontinuation worsens physical functioning, and along with depressive behaviors, this could result in further impairments in social relationships and a consequential reduced overall quality of life [10]. HIV/AIDS and depression exhibit overlapping symptoms such as fatigue, sleep problems, bereavement, loss of appetite and cognitive impairment [5] and this has led to the under diagnosis of MDD in HIV and in many instances HIV too goes undiagnosed. This misdiagnosis of MDD in PLWHIV has been under-recognized [10] and is often erroneously attributed to psychological distress related to the HIV diagnosis.

Several studies have investigated socio-demographic, psychosocial and clinical risk factors of MDD in HIV but there remains a critical gap in understanding its genetics which is crucial for unraveling the biological mechanisms underlying the disorder [11]. This has been addressed through this Genome Wide Association Study (GWAS) to identify genetic loci that predispose PLWHIV in Masaka and Entebbe districts, Uganda to depression.

Genome-wide approaches are now widely used in medical research; however, few such studies have been conducted in Low and Middle Income Countries (LMICs) [12]. Recent GWAS meta-analyses have identified common genetic variants that are associated with MDD [3, 13, 14], but these have mainly been done on individuals of European ancestry and for those that have included Africans [14], they are African-admixed participants in the United States and United Kingdom and may not represent the continental Africans [15]. These populations often have genetic contributions from Europeans, Native Americans, or other groups, which can skew results in studies aimed at understanding genetic diversity or disease susceptibility specifically in populations of purely African descent [16]. This admixture can obscure signals of genetic variants that are more prevalent or functionally

important in continental African populations, leading to inaccurate or incomplete conclusions about African genetics. There is a need to understand the genetics of depression not only in European populations, but also in other ethnic groups since several studies have shown that psychiatric disorders may vary among race and ethnic groups [17] calling for more research on the etiology of MDD in understudied populations, especially Africans. African populations harbor the greatest genetic diversity globally, making them a rich resource for identifying rare and common genetic variants that may not exist in European or other populations [18].

Therefore, there is a dearth of studies on the genetic factors that underlie MDD among populations of African ancestry as these may facilitate the identification of genetic markers that could represent novel mechanistic pathways that could enable the development of therapeutics that are more efficacious and have minimal side effects. Secondly, no GWAS has ever been undertaken for MDD among PLWHIV, yet MDD is a significant cause of disability in HIV/AIDS [6, 10, 19]. Additionally, findings from such studies can be shared with larger study consortia such as the Psychiatric Genomics Consortium to contribute to wider research efforts towards understanding the genetic risk for MDD.

2. Materials and Methods

This case-control genetics study used samples from an EDCTP-funded project (grant number: Ta. K. 40,200.01) that explored whether mental health factors should be routinely controlled for in HIV/AIDS clinical trials in Africa. The EDCTP study focused on mental health among HIV-positive patients at TASO clinics in Entebbe and Masaka, Uganda. These locations were selected to assess the target population (PLWHIV) and to include both urban (Entebbe) and rural (Masaka) participants.

2.1 Identifying samples for the genetics sub-study and their description

In the primary EDCTP-funded study, approximately 3,000 and 2,500 ART-naïve clients attended the TASO Entebbe and Masaka clinics, respectively, at that time. A random sample of 550 ART-naïve patients from each clinic was selected, giving a combined sample of 1,100 participants for the main study. Out of these, 750 samples were randomly selected to be used for a nested genetics sub-study [20], and this GWAS study used 282 of the 750 samples. Patient data for these 282 individuals (139 with major depressive disorder (MDD) and 143 age matched controls) were retrieved from the biobank at the Medical Research Council /Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine Uganda Research Unit MRC/UVRI & LSHTM Uganda Research Unit. Only cases meeting DSM-IV criteria for MDD were included. Participants were HIV-positive, ART-naïve, at least 18 years old, conversant in Luganda, and had given written informed consent. Those who were too sick or had defaulted on their most recent clinic visits were excluded. Participants were assessed for major depressive disorder (MDD) at baseline, 6 months, and 12 months. MDD diagnosis was conducted using the MDD module of the Mini International Neuropsychiatric Interview-Plus (M.I.N.I. Plus) [21], a structured tool for diagnosing mental disorders according to DSM-IV, administered by psychiatric nurses.

2.2 DNA extraction

DNA was extracted from 3 mL of whole blood samples using the Qiamp Mini DNA Extraction Kit (Qiagen, Germany), following manufacturer's instructions and suspended in 50 µL of nuclease-free water to prevent degradation. The quality and quantity of the DNA were assessed using spectrophotometer absorbance ratios (260:280) and stored at -20°C until shipment. The samples were then shipped to the Centre for Proteomics and Genomic Research (CPGR) in South Africa for genotyping. Genotyping was performed using the Infinium LCG Assay protocol with the custom H3Africa v2 array BeadChip (Illumina product code: ILL-15056944), covering over 2 million SNPs across various African population groups.

2.3 Data management and statistical analysis

2.3.1 Raw Data Quality Control

The raw data in Intensity Data (IDAT) format was assessed using GenomeStudio 2.0 software [22]. To load the IDAT files, a sample sheet (containing sample information and indexes) and a manifest file (describing SNPs or probes on the BeadChip) were required. Samples with call rates of 0.95 or higher passed the quality control (QC). The GenTrain score, which measures the quality of SNP calling (ranging from 0.0 to 1.0), was also used for data quality assessment, with reliable scores between 0.5 and 1.0. Only SNPs that met these criteria were retained for further analysis.

GenomeStudio was also to convert the IDAT files into PLINK-compatible file formats (.bed, .bim, .fam), which contain genotype information, SNP details, and participant data. These files were shared with the mental health unit at MRC/UVRI Uganda and Makerere University for GWAS.

2.3.2 Genome-wide association analysis

The GWAS was performed in four steps: data cleaning and quality control, data generation, genome-wide association analysis, and post-analytic visualization [23]. Each step of the GWAS process plays a critical role in ensuring the reliability and interpretability of results [13, 23]. Data cleaning and quality control (QC) involve assessing the raw genetic data for missingness, duplicates, and potential errors such as sample contamination or population stratification. This step ensures that only high-quality data are included in subsequent analyses [13, 23]. Data generation included SNP imputation from the Sanger imputation server [24] and genotyping using high-throughput platforms to identify single nucleotide polymorphisms (SNPs) across the genome [13, 23, 24]. The genome-wide association analysis examines the relationship between each SNP and the phenotype of interest (MDD in this case), leveraging statistical models to identify loci associated with the disorder [25]. Finally, post-analytic visualization uses plots such as Manhattan and quantile-quantile (Q-Q) plots to present findings, highlight significant associations, and assess the overall quality of the analysis.

Quality Control and Data Cleaning: The genotype data was filtered at both the SNP and sample levels using criteria such as SNP-level missingness, individual-level missingness, minor allele frequency (MAF), Hardy-Weinberg equilibrium (HWE), SNP

heterozygosity, relatedness, and population stratification using Plink 1.9 [26]. SNPs with a call rate < 0.95 and MAF < 0.01 were removed. The Hardy-Weinberg test filtered out SNPs with a p-value ≤ $1x10^{-6}$. Principal component analysis (PCA) was used to account for population substructure by including the first ten principal components in the GWAS model. At the sample level, a call rate cut-off of 0.95 was used to exclude samples that had low call rates that could have been a result of poor quality DNA. Individuals with an inbreeding coefficient < 0.1 or high relatedness (identity by descent, IBD, pi_ha ≥ 0.1) were filtered out. These thresholds were selected with guidance from [13]. Additionally, the genomic inflation factor (λGC) was calculated to assess potential residual population stratification. These measures, combined with stringent filtering criteria, reduce bias and improve the reproducibility of findings.

Data Imputation: After data cleaning and quality control, missing or untyped SNPs were imputed using reference data from the African Genome Resource via the Sanger Imputation Server [24], ensuring accurate imputation of population-specific variants. Imputed SNPs were subjected to QC, filtering for MAF ≥ 0.01, linkage disequilibrium (LD, $r^2 \geq 0.5$), and removing SNPs with low imputation accuracy (info score < 0.7). The imputation process ensured comprehensive genome coverage, allowing for the inclusion of variants that may be specific to African populations and otherwise absent from genotyping arrays.

Genome-Wide Association Analysis: The GWAS was performed using GEMMA [25] software, which applied a univariate linear mixed model (LMM) to test associations between SNPs (both genotyped and imputed) and MDD. The model adjusted for age, gender, and principal components. The genome-wide significance threshold was set at $p \leq 5\times10^{-8}$. A suggestive threshold ($p \leq 1\times10^{-6}$) was used to identify SNPs that exhibited a trend towards significant association with MDD. This threshold was based on the fact that it is a less stringent cutoff which balances exploratory power with statistical rigor, highlighting loci that might merit further investigation in replication studies or follow-up research [27].

Post-Analytic Visualization: Post-analytic visualization involved Manhattan plots to show significant SNPs by chromosome location, quantile-quantile (Q-Q) plots to compare expected and observed p-value distributions, and LD heatmaps to visualize patterns of linkage disequilibrium between significant SNPs. These plots were created using R [28] packages; ggplot2 [29] and qqman [30].

2.3.3 Functional mapping and annotation

Functional mapping and annotation (FUMA) [31] software was used to analyze GWAS data. FUMA, a web-based tool, works in two core processes: SNP2GENE and GENE2FUNC, using GWAS summary statistics as input. SNP2GENE identified functional SNPs and genes, outputting tables with significant SNPs, genomic risk loci, SNP positions, and mapped genes. It also generated Manhattan, QQ, and interactive regional plots, similar to those created in R-studio. GENE2FUNC provided outputs such as gene expression heatmaps, enrichment of differentially expressed gene (DEG) sets in specific tissues, and links to external biological databases for further information on input genes. These processes helped prioritize and visualize significant genetic associations.

3. Results

3.1 Participant demographics

The parent study [8], reported detailed participant demographics, indicating that the majority of respondents were female. The mean age of the GWAS participants (282 persons) was 35.1 years with a standard deviation (SD) of 9.3. One hundred seventy nine participants were drawn from the rural area, Masaka TASO clinic and one hundred four participants were from semi- urban area, Entebbe TASO clinics.

3.2 Quality control of sample DNA quality

After proper age matching, 282 DNA samples were sent to CPGR for genotyping. CPGR conducted quality checks on the DNA, finding that 281 samples had an A260/280 absorbance ratio within the expected range of 1.8–2.1, with one sample slightly below. As the A260/A280 ratio was not deemed critical to the overall results according to validated Illumina-Infinium® protocols, all samples were included in genotyping.

3.3 Quality control of raw reads

The quality of the raw data from the genotyping array was assessed using GenomeStudio 2.0 software and is summarized in Table 1.

Table 1. Table showing quality metrics used for cleaning raw reads after genotyping from 282 samples.

Quality metric		
Sample call rate (> 0.95)	Number of passed samples	278
	Number of failed samples	4
	Percentage success	98.5%
GenTrain score (0.0 – 1.0)	Number of SNPs with score => 0.5	2,247,439
	Number of failed SNPs	24,064
	Percentage success	98.4%

Sample Call Rate: The percentage of successful genotype calls for each sample, meaning 95% of calls per sample must be successful to pass quality control).
GenTrain Score: A score assessing the quality of SNP genotype clusters, where scores closer to 1.0 indicate better clustering and reliability of the SNP call. A score ≥ 0.5 provides an optimal trade-off between sensitivity, including SNPs likely associated with the disease and specificity, avoiding false positives.

3.4 Genome-wide association results

The quality filtered genotype data consisted of 282 participant samples with 2,247,439 SNPs. QC filtering based on a 95% call rate removed 20,147 SNPs. Filtering for minor allele frequency (MAF) using a 0.01 cut-off removed 1,348,601 SNPs. Hardy-Weinberg equilibrium (HWE) filtering at a cut-off of $1x10^{-6}$ eliminated 598 SNPs, and 452,748 SNPs were removed due to high linkage disequilibrium (LD) above r^2 >0.5. At the sample level, one sample was removed for heterozygosity (cut-off 0.1), and two

for close relatedness. After QC, the final dataset included 275 participants (134 MDD cases, 141 controls) with 409,768 SNPs. Imputation using the African Genome Resource via the Sanger imputation service generated over 89 million SNPs, which were also subjected to QC.

3.5 Principal component analysis

Principal component analysis (PCA) of the genotype data, shown in Figure 1, indicated that the samples had similar ancestry, except for the 3 samples, as highlighted in the PCA plot. The first two principal components together, explained 29.05% of the variance.

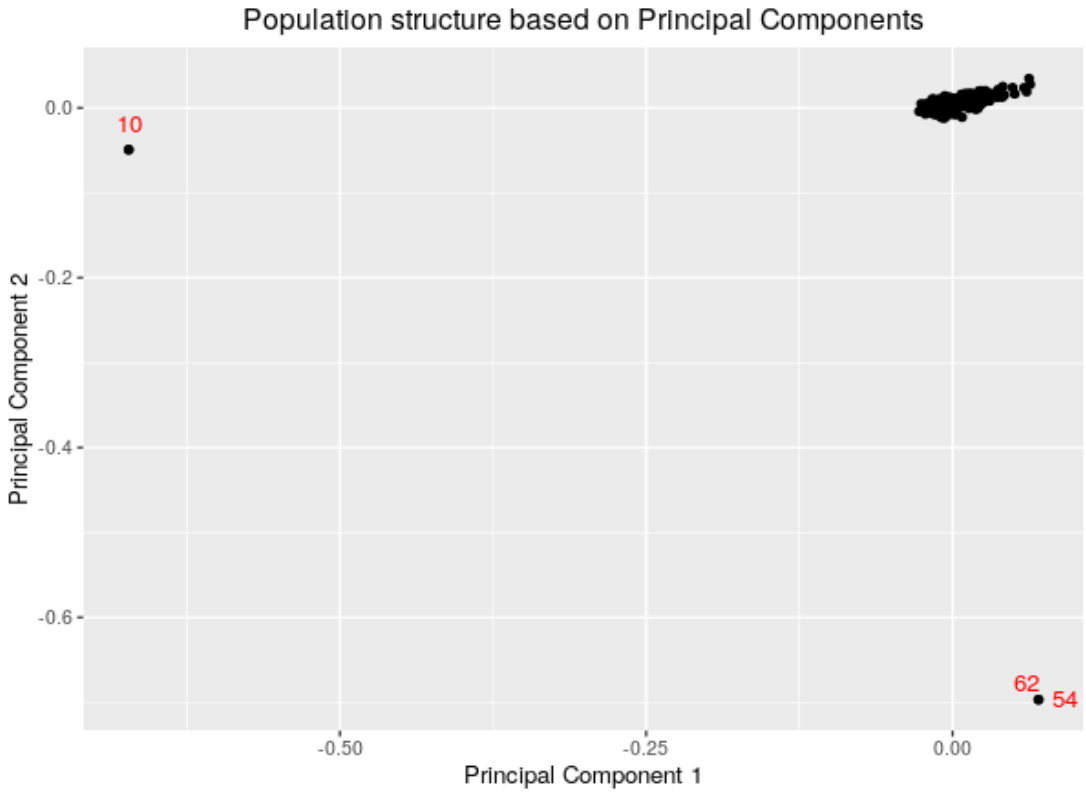

Figure 1. A summary of the first two principal components of study participants to illustrate population structure. The 3 outlier samples are identified by sample IDs highlighted in red.

3.6 Quantile plot visualization

The QQ plot in Figure 2 below shows that the observed results aligned well with what would be expected by chance, with SNPs closely following the expected slope. A slight deviation in the upper right tail suggested a possible weak association. There was no evidence of genomic inflation or bias due to population stratification, as indicated by a λ-statistic of 0.9661.

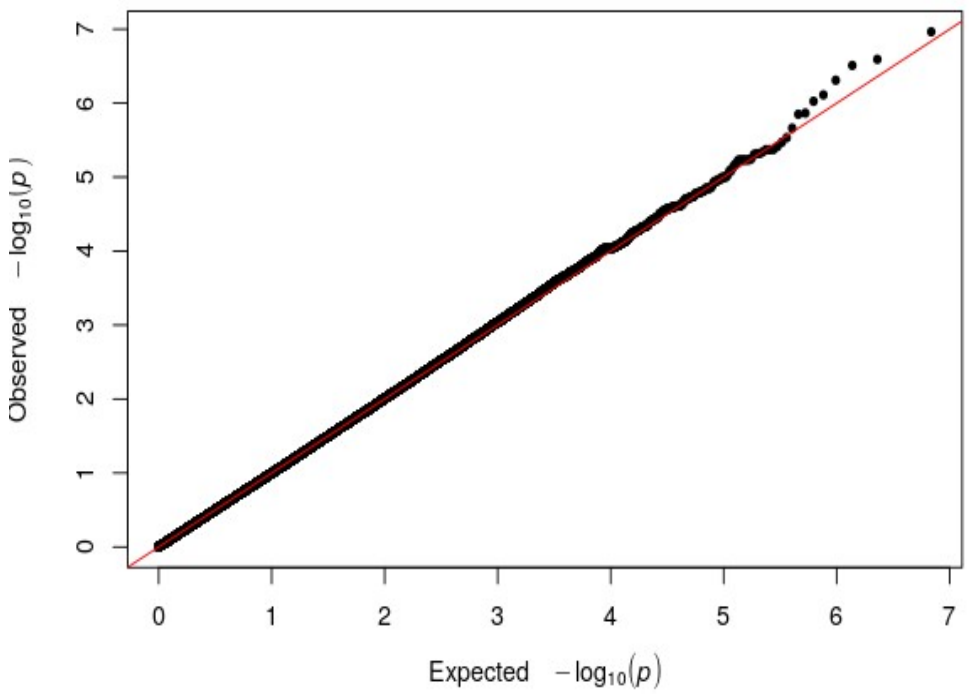

Figure 2. A Q-Q plot showing the extent to which the observed distribution of the p-values follow the expected p-values

3.7 Functional mapping and annotation

We did not find any significant SNPs at the genome-wide significance threshold ($p \leq 5\times10^{-8}$). However, at our suggestive significance level of ($p \leq 1\times10^{-6}$), six SNPs were identified as potentially associated with MDD in this cohort. Table 2 shows the first top 10 leading SNPs from the analysis that could be worth further investigating.

Table 2. A table showing the top ten leading SNPs from the analysis.

Chromosome	RS ID	Allele 1	Allele 2	Allele frequency	Beta	Standard Error	P value
2	rs4375916	G	A	0.456	0.223	0.041	1.086×10^{-7}
2	rs2348723	A	G	0.420	0.216	0.042	2.558×10^{-7}
3	rs1075618	A	G	0.264	0.255	0.049	3.093×10^{-7}
13	rs9531855	G	T	0.207	-0.271	0.053	4.906×10^{-7}
2	rs12713006	T	A	0.404	0.209	0.041	7.741×10^{-7}
2	rs6705995	G	T	0.538	0.205	0.041	9.437×10^{-7}
5	rs630979	C	G	0.685	-0.206	0.042	1.362×10^{-6}
2	rs1546470	C	G	0.413	0.204	0.041	1.415×10^{-6}

1	rs2022003	T	A	0.295	0.218	0.045	2.181×10^{-6}
2	rs55802119	C	A	0.431	-0.200	0.042	2.930×10^{-6}

Functional mapping and annotation revealed three genomic risk loci at the suggestive threshold of $p\leq1\times10^{-6}$. The Manhattan plot in Figure 1, supplementary materials, shows SNP association results at this threshold. Six SNPs reached suggestive genome-wide significance ($p\leq1\times10^{-6}$), with the most significant being rs4375916 on chromosome 2 ($p\leq1.086\times10^{-7}$). These six SNPs mapped to 132 protein-coding genes, 79 of which showed differential tissue expression in 54 tissue types. Two genes *SLITRK6 and RN7SKP224-FOXN2*—displayed genome-wide significance.

Discussion

This preliminary study explored the genetic risk of (MDD in 282 people living with HIV (PLWHIV) attending ART clinics at TASO Entebbe and TASO Masaka in Uganda. The study had a modest sample size of 282 participants on account of funding constraints.

The GWAS analysis identified six suggestively significant SNPs in three genomic risk loci associated with MDD. These SNPs mapped to 132 protein-coding genes, with 79 expressed in 54 tissue types, including the brain. The low number of risk loci, compared to larger studies in European ancestry populations [1, 3], is likely due to the small sample size (275 participants after quality control). The study had limited statistical power, with only a 5.7% chance of detecting a true effect [32]. Despite the study's low statistical power, six significant SNPs were identified using a relaxed significance threshold of $p\leq1\times10^{-6}$, which is less stringent than the commonly used threshold of $p\leq8\times10^{-5}$ for GWAS studies [27]. This adjusted threshold is often applied in underpowered studies and exploratory studies [33] to generate hypotheses that will be tested in larger studies [34], minimizing false negatives while ensuring that potential findings are not missed [35].

Of note, among the six suggestively significant SNP, while none has been previously identified as associated with MDD, have been reported in GWAS studies of other psychiatric conditions (Alzheimer's disease and alcohol dependency) – conditions that are comorbid with MDD [36, 37]. For instance the SNP rs4375916, has previously been described to have a significant association with Alzheimer's [36]. MDD and Alzheimer's have been reported to have a shared polygenic component [38, 39, 40] and also share common neurobiological abnormalities [40]. However some studies found no evidence to support a common polygenic structure for Alzheimer's and MDD [41], which was in contrast with what some studies have reported [37, 42]. The inconsistency in these findings highlights the need for larger, more comprehensive studies to establish conclusive evidence regarding genetic associations with these disorders. Individuals of African American ancestry have been reported to be at higher risk of Alzheimer's than persons of European ancestry for reasons that may include economic disparities, cardiovascular health, quality of education,

and biases in the methods used to diagnose Alzheimer's [42]. Such studies are essential as they could enable reliable precision medicine approaches in persons with considerable African ancestry. Even though some studies have reported shared etiology between MDD and Alzheimer's, we cannot confidently conclude that Africans or African Americans could be at higher risk for MDD as in Alzheimer's.

rs4375916 mapped to the RN7SKP224 - FOXN2 gene, a pseudo gene, in our study which has also been identified in Alzheimer's [36], however its proper role in Alzheimer's has not been established as yet and also no literature has been found to define its role in MDD or the brain tissue and related organs.

The SNP rs9531855, has also been previously identified in a study that conducted a meta-analysis of genetic influences on initial alcohol sensitivity [43]. This is not surprising as a number of studies have reported a prevalence of comorbidity of depression and alcohol use disorders (AUD) [43, 44]. Research indicates that higher alcohol consumption, influenced by genetics and environmental factors, increases the risk of developing major depression [45]. Alcohol use disorder and major depression each double the risk of the other [45, 46], and numerous studies show a strong association between alcohol dependence and depression [47]. Depression among those with alcohol dependence may weaken their resolve to avoid alcohol and may drive alcohol use as a way to alleviate depressive symptoms [47]. There are two possible explanations for the association between alcohol use disorders and major depression; firstly it may be that both disorders have common underlying genetic and environmental factors that jointly increase the risk of both disorders. Secondly, the two disorders may have a causal effect with each disorder increasing the risk of developing the other [46]. This SNP mapped to the *SLITRK6* gene that possesses variants which have been previously identified to have a role in bipolar GWAS [47]. Bipolar disorder and (MDD) can sometimes be mistaken for one another, even by mental health professionals, due to the presence of depressive episodes in both conditions [48, 49]. Bipolar and MDD have been identified to have lots of overlaps especially in their genetic epidemiology and molecular genetics [36, 50, 51, 52].

A limitation of this study was the small sample size resulting in an underpowered study (study power - 5.7%) increasing the likelihood of false-positive results. Small sample size is a frequent problem in studies of polygenic diseases and can result in insufficient power to detect minor contributions of one or more alleles. Similarly, small sample sizes can provide imprecise or incorrect estimates of the magnitude of the observed effects. For a highly powered GWAS for polygenic diseases, over thousands and millions of study participants are needed to have a conclusive study with no false positives associated to the discoveries [53]. Lastly, confounding may lie in the fact that most participants had comorbid psychiatric diagnoses, as in the case in most samples of psychiatric disorders. However, data on psychiatric comorbidities were not collected. Overall, the research into the genetics of MDD is limited especially in native African ancestry. As such, this study provides potentially useful preliminary data and methodological approaches to inform future investigations of this important area despite being unable to detect any significant

associations in any of our analyses at known thresholds. Findings also present plausible evidence that with a larger sample size, there is a potential to identify more independent risk loci and significant SNPs associated with MDD in an African ancestry population.

5. Conclusion

This study did not identify any genome-wide significant SNPs or genetic risk loci, most likely due to a small sample size, and consequent lack of statistical power. However, applying a genome-wide suggestive threshold of ($p \leq 1\times10^{-6}$) identified three genetic risk loci, and six suggestively genome-wide significant SNPs, three of which have not been previously been found to be associated with MDD. The results from this study also suggest that psychosocial factors and psychosomatic complaints are the major risk factors for MDD. These results indicate a high potential for novel MDD genetic risk loci discovery in African ancestry populations especially since there were suggestively significant findings obtained even with a small sample size. Therefore, large sample size studies of this kind are highly recommended within the African population as these could be avenues of novel loci discovery.

6. List of abbreviations

MDD: major depressive disorder
GWAS: genome-wide association study
YLD: years lived with disability
WHO: world health organisation
HIV: human immunodeficiency virus
PLWHIV: persons living with HIV
DSM: diagnostic and statistical manual of mental disorders
M.I.N.I: Mini International Neuropsychiatric Interview
CD4: cluster of differentiation 4
SNP: single nucleotide polymorphism
EDCTP: European and developing countries clinical trials partnership
ART: anti-retroviral therapy
TASO: The AIDS Support Organisation
MRC: Medical Research Council
UVRI: Uganda Virus Research Institute
LSHTM: London School of Hygiene & Tropical Medicine

EDTA ethylene diamine tetra acetic acid

MAF: minor allele frequency

HWE: Hardy– Weinberg equilibrium

PC: principal component

Declarations

Ethical Approval and consent to participate

This study was approved by the School of Biomedical Sciences Research Ethics Committee, Makerere University, reference number: SBS-2021-48. All participants provided written informed consent in the edctp study by Eugene Kinyanda and the genetics study by Allan Kalungi. This study did not have to seek consent to participate. The research was conducted in accordance with the Declaration of Helsinki.

Clinical trial number

Not applicable

Consent for publication

Not applicable.

Availability of data and materials

The dataset presented in this article are currently not publicly available as further analysis is ongoing to address additional research questions. The datasets can be made available from the corresponding author on reasonable request.

Competing interests

The authors declare no conflicts of interests.

Funding

Research herein was partially funded from the DSI/SAMRC Africa Health Research, Development and Innovation Program and the Brain Health Project D43NS118560 of Makerere University. Allan Kalungi is a Wellcome Early Career Fellow grant number: 227053/Z/23/Z and was also supported by the 2020 Brain and Behaviour Research Foundation (formerly NARSAD) young investigator grant (grant No. 29610)

Authors'Contributions

Conceptualization, Olga.N.T., Allan.K., Eugene.K., Sian.H., and Soraya.S; methodology, Olga.N.T.; software, Olga.N.T.; validation, OlgaN.T., Ronald.G., Sian.H., Martha.S., Catherine.N., Mark.K., Martin.K., Eegene.K., Allan.K., Soraya.S., and Elly.K.; formal analysis, Olga.N.T.; writing—original draft preparation, OlgaN.T.; supervision, Ronald.G., Sian.H., Martha.S.,

Mark.K., Marttin.K., Eugene.K., Allan.K., and Elly.K.; funding acquisition, Olga.N.T, Sian.H, Soraya.S, Allan.K and Eugene.K, writing—original draft preparation,; OlgaN.T., Ronald.G., Sian.H., Martha.S., Catherine.N., Mark.K., Martin.K., Eegene.K., Allan.K., Soraya.S., and Elly.K All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We acknowledge BioTeam for providing computational resources that were instrumental in the analysis of the data presented in this study. Research reported in this publication was supported by the South African Medical Research Council (SAMRC) Genomics of Brain Disorders (GBD) Unit. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the DSI/SAMRC. Gratitude to the African Centre of Excellence, Infectious Diseases Institute and the Department of immunology and Molecular Biology, College of Health Science, Makerere University and the Brain Health Project D43NS118560, for training and mentorship received from them to enable complete this research.

Authors' information

Authors and affiliations

Olga Nsangi Tendo: Department of Immunology and Molecular Biology, Makerere University, P.O.Box 7072, Kampala, Uganda and African Centers of Excellence in Bioinformatics and Data Intensive Sciences, infectious Diseases Institute, Makerere University, Kampala, Uganda

Ronald Galiwango: Department of Immunology and Molecular Biology, Makerere University, P.O.Box 7072, Kampala, Uganda and African Centers of Excellence in Bioinformatics and Data Intensive Sciences, infectious Diseases Institute, Makerere University, Kampala, Uganda

Eugene Kinyanda: Mental Health Unit of Medical Research Council/Uganda Virus Research Institute (MRC/UVRI) and London School of Hygiene & Tropical Medicine(LSHTM) Uganda Research Unit, Uganda

Martha Sajatovic: School of Medicine, Case Western Reserve University School,United States of America

Kaddumukasa Mark: Department of Medicine, Makerere University, P.O.Box 7072, Kampala, Uganda

Kaddumukasa Martin: Department of Medicine, Makerere University, P.O.Box 7072, Kampala, Uganda

Elly Katabira: Department of Medicine, Makerere University, P.O.Box 7072, Kampala, Uganda

Nabbumba Catherine: Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Soraya Seedat: Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa and South African Medical Research Council/Stellenbosch University Genomics of Brain Disorders Unit, Department of

Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Sian Megan Joanna Hemmings: Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa and South African Medical Research Council/Stellenbosch University Genomics of Brain Disorders Unit, Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Allan Kalungi: Medical Research Council/Uganda Virus Research Institute (MRC/UVRI) and London School of Hygiene & Tropical Medicine (LSHTM) Uganda Research Unit, Uganda.

References

[1] J. Daly *et al.*, "A mega-analysis of genome-wide association studies for major depressive disorder.," *Mol. Psychiatry*, vol. 18, no. 4, Art. no. 4, 2013.

[2] "WHO-MSD-MER-2017.2-eng.pdf."

[3] D. M. Howard *et al.*, "Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions," *Nat. Neurosci.*, vol. 22, no. 3, Art. no. 3, Mar. 2019, doi: 10.1038/s41593-018-0326-7.

[4] E. Kinyanda, S. Hoskins, J. Nakku, S. Nawaz, and V. Patel, "Prevalence and risk factors of major depressive disorder in HIV/AIDS as seen in semi-urban Entebbe district, Uganda," *BMC Psychiatry*, vol. 11, p. 205, Dec. 2011, doi: 10.1186/1471-244X-11-205.

[5] C. Bernard, F. Dabis, and N. de Rekeneire, "Prevalence and factors associated with depression in people living with HIV in sub-Saharan Africa: A systematic review and meta-analysis," *PLOS ONE*, vol. 12, no. 8, p. e0181960, Aug. 2017, doi: 10.1371/journal.pone.0181960.

[6] B. X. Tran *et al.*, "Depression among Patients with HIV/AIDS: Research Development and Effective Interventions (GAPRESEARCH)," *Int. J. Environ. Res. Public. Health*, vol. 16, no. 10, p. 1772, May 2019, doi: 10.3390/ijerph16101772.

[7] D. M. Mulqueeny, S. M. Nkabini, and M. H. Pokiya, "Mapping evidence of depression in HIV-seropositive MSM in sub-Saharan Africa: a scoping review protocol," *Syst. Rev.*, vol. 10, p. 50, Feb. 2021, doi: 10.1186/s13643-021-01604-w.

[8] E. Kinyanda *et al.*, "Major depressive disorder and suicidality in early HIV infection and its association with risk factors and negative outcomes as seen in semi-urban and rural Uganda," *J. Affect. Disord.*, vol. 212, pp. 117–127, Apr. 2017, doi: 10.1016/j.jad.2017.01.033.

[9] E. Kinyanda *et al.*, "Major Depressive Disorder: Longitudinal Analysis of Impact on Clinical and Behavioral Outcomes in Uganda," *JAIDS J. Acquir. Immune Defic. Syndr.*, vol. 78, no. 2, p. 136, Jun. 2018, doi: 10.1097/QAI.0000000000001647.

[10] M. Abas, G.-C. Ali, E. Nakimuli-Mpungu, and D. Chibanda, "Depression in people living with HIV in sub-Saharan Africa: time to act," *Trop. Med. Int. Health*, vol. 19, no. 12, pp. 1392–1396, 2014, doi: 10.1111/tmi.12382.

[11] M. C. Greene *et al.*, "The epidemiology of psychiatric disorders in Africa: a scoping review," *Lancet Psychiatry*, vol. 8, no. 8, pp. 717–731, Aug. 2021, doi: 10.1016/S2215-0366(21)00009-2.

[12] H. Shen, B. Gelaye, H. Huang, M. B. Rondon, S. Sanchez, and L. E. Duncan, "Polygenic prediction and GWAS of depression, PTSD, and suicidal ideation/self-harm in a Peruvian cohort," *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.*, vol. 45, no. 10, pp. 1595–1602, Sep. 2020, doi: 10.1038/s41386-020-0603-5.

[13] A. T. Marees *et al.*, "A tutorial on conducting genome-wide association studies: Quality control and statistical analysis," *Int. J. Methods Psychiatr. Res.*, vol. 27, no. 2, p. e1608, Feb. 2018, doi: 10.1002/mpr.1608.

[14] X. Meng *et al.*, "Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference," *Nat. Genet.*, vol. 56, no. 2, pp. 222–233, Feb. 2024, doi: 10.1038/s41588-023-01596-4.

[15] C. Jonson *et al.*, "Assessing the lack of diversity in genetics research across neurodegenerative diseases: a systematic review of the GWAS Catalog and literature," *medRxiv*, p. 2024.01.08.24301007, Jan. 2024, doi: 10.1101/2024.01.08.24301007.

[16] L. Vicuña, "Genetic associations with disease in populations with Indigenous American ancestries," *Genet. Mol. Biol.*, vol. 47, p. e20230024, Sep. 2024, doi: 10.1590/1678-4685-GMB-2023-0024.

[17] D. C. Watkins, S. Assari, and V. Johnson-Lawrence, "Race and Ethnic Group Differences in Comorbid Major Depressive Disorder, Generalized Anxiety Disorder, and Chronic Medical Conditions," *J. Racial Ethn. Health Disparities*, vol. 2, no. 3, pp. 385–394, Sep. 2015, doi: 10.1007/s40615-015-0085-z.

[18] S. Dalvie *et al.*, "Large Scale Genetic Research on Neuropsychiatric Disorders in African Populations is Needed," *EBioMedicine*, vol. 2, no. 10, p. 1259, Oct. 2015, doi: 10.1016/j.ebiom.2015.10.002.

[19] K. O. Bankole *et al.*, "Psychological complications associated with HIV/AIDS infection among children in South-South Nigeria, sub-Saharan Africa," *Cogent Med.*, vol. 4, no. 1, p. 1372869, Jan. 2017, doi: 10.1080/2331205X.2017.1372869.

[20] A. Kalungi *et al.*, "Association between serotonin transporter gene polymorphisms and increased suicidal risk among HIV positive patients in Uganda," *BMC Genet.*, vol. 18, no. 1, p. 71, Jul. 2017, doi: 10.1186/s12863-017-0538-y.

[21] I. M. van Vliet and E. de Beurs, "[The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders]," *Tijdschr. Voor Psychiatr.*, vol. 49, no. 6, pp. 393–397, 2007.

[22] "GenomeStudio Software Downloads." Accessed: Oct. 27, 2024. [Online]. Available: https://support.illumina.com/array/array_software/genomestudio/downloads.html

[23] E. Reed, S. Nunez, D. Kulp, J. Qian, M. P. Reilly, and A. S. Foulkes, "A guide to genome-wide association analysis and post-analytic interrogation," *Stat. Med.*, vol. 34, no. 28, pp. 3769–3792, Dec. 2015, doi: 10.1002/sim.6605.

[24] "Sanger Imputation Service." Accessed: Aug. 01, 2023. [Online]. Available: https://www.sanger.ac.uk/tool/sanger-imputation-service/

[25] X. Zhou and M. Stephens, "Genome-wide Efficient Mixed Model Analysis for Association Studies," *Nat. Genet.*, vol. 44, no. 7, pp. 821–824, Jun. 2012, doi: 10.1038/ng.2310.

[26] S. Purcell *et al.*, "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses," *Am. J. Hum. Genet.*, vol. 81, no. 3, p. 559, Jul. 2007, doi: 10.1086/519795.

[27] Z. Chen, M. Boehnke, X. Wen, and B. Mukherjee, "Revisiting the genome-wide significance threshold for common variant GWAS," *G3 GenesGenomesGenetics*, vol. 11, no. 2, p. jkaa056, Feb. 2021, doi: 10.1093/g3journal/jkaa056.

[28] "R: The R Project for Statistical Computing." Accessed: Nov. 19, 2024. [Online]. Available: https://www.r-project.org/

[29] "Create Elegant Data Visualisations Using the Grammar of Graphics." Accessed: Nov. 14, 2024. [Online]. Available: https://ggplot2.tidyverse.org/

[30] "Journal of Open Source Software: qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots." Accessed: Nov. 14, 2024. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.00731

[31] K. Watanabe, E. Taskesen, A. van Bochoven, and D. Posthuma, "Functional mapping and annotation of genetic associations with FUMA," *Nat. Commun.*, vol. 8, no. 1, p. 1826, Nov. 2017, doi: 10.1038/s41467-017-01261-5.

[32] J. L. Johnson and G. R. Abecasis, "GAS Power Calculator: web-based power calculator for genetic association studies," Jul. 17, 2017, *bioRxiv*. doi: 10.1101/164343.

[33] A. Buniello *et al.*, "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1005–D1012, Jan. 2019, doi: 10.1093/nar/gky1120.

[34] Y. Guo *et al.*, "Genome-wide detection of genetic markers associated with growth and fatness in four pig populations using four approaches," *Genet. Sel. Evol. GSE*, vol. 49, p. 21, Feb. 2017, doi: 10.1186/s12711-017-0295-4.

[35] F. Dudbridge and P. Newcombe, "Replication and Meta-analysis of Genome-Wide Association Studies," in *Handbook of Statistical Genomics*, John Wiley & Sons, Ltd, 2019, pp. 631–650. doi: 10.1002/9781119487845.ch22.

[36] C. Gouveia, E. Gibbons, N. Dehghani, J. Eapen, R. Guerreiro, and J. Bras, "Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank," *Sci. Rep.*, vol. 12, p. 8404, May 2022, doi: 10.1038/s41598-022-12391-2.

[37] A. C. Edwards *et al.*, "Meta-Analysis of Genetic Influences on Initial Alcohol Sensitivity," *Alcohol. Clin. Exp. Res.*, vol. 42, no. 12, pp. 2349–2359, Dec. 2018, doi: 10.1111/acer.13896.

[38] M. W. Lutz, D. Sprague, J. Barrera, and O. Chiba-Falek, "Shared genetic etiology underlying Alzheimer's disease and major depressive disorder," *Transl. Psychiatry*, vol. 10, no. 1, Art. no. 1, Mar. 2020, doi: 10.1038/s41398-020-0769-y.

[39] E. Kitzlerová *et al.*, "Interactions Among Polymorphisms of Susceptibility Loci for Alzheimer's Disease or Depressive Disorder," *Med. Sci. Monit.*, vol. 24, pp. 2599–2619, Apr. 2018, doi: 10.12659/MSM.907202.

[40] M.-S. Mendes-Silva, Ana Paula Ana Paula, "Shared Biologic Pathways Between Alzheimer Disease and Major Depression:

A Systematic Review of MicroRNA Expression Studies - PubMed." Accessed: Jul. 20, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27591915/

[41] J. Gibson *et al.*, "Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data," *Transl. Psychiatry*, vol. 7, no. 4, Art. no. 4, Apr. 2017, doi: 10.1038/tp.2017.49.

[42] M. W. Logue, S. Dasgupta, and L. A. Farrer, "Genetics of Alzheimer's Disease in the African American Population," *J. Clin. Med.*, vol. 12, no. 16, p. 5189, Aug. 2023, doi: 10.3390/jcm12165189.

[43] R. C. Kessler, R. M. Crum, L. A. Warner, C. B. Nelson, J. Schulenberg, and J. C. Anthony, "Lifetime Co-occurrence of DSM-III-R Alcohol Abuse and Dependence With Other Psychiatric Disorders in the National Comorbidity Survey," *Arch. Gen. Psychiatry*, vol. 54, no. 4, pp. 313–321, Apr. 1997, doi: 10.1001/archpsyc.1997.01830160031005.

[44] H. E. Ross, "DSM-III-R alcohol abuse and dependence and psychiatric comorbidity in Ontario: results from the Mental Health Supplement to the Ontario Health Survey," *Drug Alcohol Depend.*, vol. 39, no. 2, pp. 111–128, Aug. 1995, doi: 10.1016/0376-8716(95)01150-W.

[45] J. M. Boden and D. M. Fergusson, "Alcohol and depression," *Addiction*, vol. 106, no. 5, pp. 906–914, 2011, doi: 10.1111/j.1360-0443.2010.03351.x.

[46] M. W. Kuria *et al.*, "The Association between Alcohol Dependence and Depression before and after Treatment for Alcohol Dependence," *ISRN Psychiatry*, vol. 2012, p. 482802, Jan. 2012, doi: 10.5402/2012/482802.

[47] T. A. Greenwood, Bipolar Genome Study (BiGS) Consortium, and J. R. Kelsoe, "Genome-wide association study of irritable vs. elated mania suggests genetic differences between clinical subtypes of bipolar disorder," *PloS One*, vol. 8, no. 1, p. e53804, 2013, doi: 10.1371/journal.pone.0053804.

[48] "Bipolar Disorder And Depression: Understanding The Difference." Accessed: Jul. 14, 2023. [Online]. Available: https://www.psycom.net/depression/bipolar-depression

[49] L. Tondo, G. H. Vázquez, M. Pinna, P. A. Vaccotto, and R. J. Baldessarini, "Characteristics of depressive and bipolar disorder patients with mixed features," *Acta Psychiatr. Scand.*, vol. 138, no. 3, pp. 243–252, Sep. 2018, doi: 10.1111/acps.12911.

[50] "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis," *The Lancet*, vol. 381, no. 9875, pp. 1371–1379, Apr. 2013, doi: 10.1016/S0140-6736(12)62129-1.

[51] Q. Yang, Q. Xing, Q. Yang, and Y. Gong, "Classification for psychiatric disorders including schizophrenia, bipolar disorder, and major depressive disorder using machine learning," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 5054–5064, Jan. 2022, doi: 10.1016/j.csbj.2022.09.014.

[52] A. T. Amare *et al.*, "Bivariate genome-wide association analyses of the broad depression phenotype combined with major

490 depressive disorder, bipolar disorder or schizophrenia reveal eight novel genetic loci for depression," *Mol. Psychiatry*, vol.
491 25, no. 7, pp. 1420–1429, Jul. 2020, doi: 10.1038/s41380-018-0336-6.
492 [53] E. P. Hong and J. W. Park, "Sample size and statistical power calculation in genetic association studies," *Genomics Inform.*,
493 vol. 10, no. 2, pp. 117–122, Jun. 2012, doi: 10.5808/GI.2012.10.2.117.