Complex Psychiatry

Research Article

Complex Psychiatry 2021;7:60–70
DOI: 10.1159/000518926

Received: February 15, 2021
Accepted: August 3, 2021
Published online: August 24, 2021

Neuropsychiatric Genetics of Psychosis in the Mexican Population: A Genome-Wide Association Study Protocol for Schizophrenia, Schizoaffective, and Bipolar Disorder Patients and Controls

Beatriz Camarena[a] Elizabeth G. Atkinson[b, c, d] Mark Baker[b] Claudia Becerra-Palars[e] Lori B. Chibnik[b, c] Raúl Escamilla-Orozco[e] Joanna Jiménez-Pavón[e] Zan Koenig[f] Carla Márquez-Luna[g] Alicia R. Martin[b, c, d] Ingrid Pamela Morales-Cedillo[a] Ana Maria Olivares[b] Hiram Ortega-Ortiz[e] Alejandra Monserrat Rodriguez-Ramírez[a] Ricardo Saracco-Alvarez[h] Rebecca E. Basaldua[f] Brena F. Sena[b, f] Karestan C. Koenen[b, f]

[a]Pharmacogenetics Department, National Institute of Psychiatry Ramón de la Fuente Muñiz, Mexico City, Mexico; [b]Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Research, Cambridge, MA, USA; [c]Analytical and Translational Genetics Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; [d]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; [e]National Institute of Psychiatry Ramón de la Fuente Muñiz, Clinical Services Direction, Mexico City, Mexico; [f]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [g]Charles R. Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [h]National Institute of Psychiatry Ramon de la Fuente Muñiz, Clinical Research Sub-direction, Mexico City, Mexico

Keywords
Neuropsychiatric genetics · Genome-wide association studies · Psychosis · Schizophrenia · Bipolar disorder · Mexico

Abstract
No large-scale genome-wide association studies (GWASs) of psychosis have been conducted in Mexico or Latin America to date. Schizophrenia and bipolar disorder in particular have been found to be highly heritable and genetically influenced. However, understanding of the biological basis of psychosis in Latin American populations is limited as previous genomic studies have almost exclusively relied on participants of Northern European ancestry. With the goal of expanding knowledge on the genomic basis of psychotic disorders within the Mexican population, the National Institute of Psychiatry Ramón de la Fuente Muñiz (INPRFM), the Harvard T.H. Chan School of Public Health, and the Broad Institute's Stanley Center for Psychiatric Research launched the Neuropsychiatric Genetics Research of Psychosis in Mexican Populations (NeuroMex) project to collect and analyze case-control psychosis samples from 5 states across Mexico. This article describes the planned sample collection and GWAS protocol for the NeuroMex study. The 4-year study will span from April 2018 to 2022 and aims to recruit 9,208 participants: 4,604 cases and 4,604 controls. Study sites across Mexico were selected to ensure collected samples capture the genomic diversity within the Mexican population. Blood samples and phenotypic data will be collected during the participant interview process and will contribute to the development of a local biobank in Mexico. DNA extraction will be done locally and genetic analysis will take place at the Broad Institute in Cambridge, MA. We will collect extensive

karger@karger.com
www.karger.com/cxp

Correspondence to:
Beatriz Camarena, camare@imp.edu.mx
Elizabeth G. Atkinson, eatkinso@broadinstitute.org

Karger

phenotypic information using several clinical scales. All study materials including phenotypic instruments utilized are openly available in Spanish and English. The described study represents a long-term collaboration of a number of institutions from across Mexico and the Boston area, including clinical psychiatrists, clinical researchers, computational biologists, and managers at the 3 collaborating institutions. The development of relevant data management, quality assurance, and analysis plans are the primary considerations in this protocol article. Extensive management and analysis processes were developed for both the phenotypic and genetic data collected. Capacity building, partnerships, and training between and among the collaborating institutions are intrinsic components to this study and its long-term success.

Introduction

Schizophrenia (SCZ) and bipolar disorder (BD) are highly heritable and polygenic neuropsychiatric conditions. BD overlaps considerably with SCZ in phenomenological expression [1], cognitive functioning [2], genetic liability [3, 4], and neuroanatomical correlates [5]. A continuum model as a psychosis spectrum has been proposed for SCZ, schizoaffective disorder (SAD), and BD [6]. Psychotic symptoms have been considered a marker for more severe psychopathology affecting the course and treatment of the disorder [7]. Family studies support overlapping heritability for SCZ, SAD, and BD [8]. A population-based study reported that first-degree relatives of patients with SCZ or BD were at increased risk of these 2 disorders [9, 10]. In addition, genome-wide association studies (GWASs) have identified genetic variants involved in both SCZ and BD, suggesting that biological pathways are shared and supporting a genetic overlap [11, 12]. A GWAS in the largest sample of patients with SCZ identified 108 significant genomic loci [13]. However, most of the samples included as part of this analysis are of European ancestry. An additional study analyzed these 108 LD-independent SCZ-associated variants among 5 major ethnic populations (East Asians, Europeans, Americans, Africans, and South Asians) finding genetic heterogeneities across these populations [14]. Recently, the largest GWAS in BD identified 30 significant loci [15]. Interestingly, a variable polygenic effect was observed between BD subtypes and in the analysis between cohorts, underscoring the importance of including a systematic clinical assessment between cases and controls [15]. In addition, genomic studies in different populations and ethnicities across the world are sorely needed to better understand the molecular etiology and mechanisms involved in psychosis in SCZ and BD within and across diverse populations [16, 17].

Over the past few years, there have been several new genetic studies of psychotic disorders in Mexico. A scan of ~400 microsatellite markers for SCZ and SAD in 99 families of Mexican and Central American ancestry identified linkage in 3 chromosomal regions that had been previously reported in other populations [18]. A genome-wide nonparametric linkage scan reported 3 loci in a sample of Mexican and Central American families with BD and SAD, bipolar type [19]. Also, a GWAS in the Latino BD cohort identified gene variants previously associated with SCZ and BD in other populations, providing evidence of shared genetic liability between the 2 disorders [20].

The complete identification of promising genes in the Latin American population has not yet been accomplished [21]. In published GWASs, Hispanic and Latin American populations represent only 0.54% of participants in GWASs who are not of European ancestry [22].

Hispanic and Latino individuals comprise a broad ethnic group within a genetically diverse continuum of populations with variations in admixture proportions from Native American, African, and European ancestries. This genetic diversity, interacting with a range of cultural and environmental factors, may have a differential impact in the susceptibility to metabolic and psychiatry disorders [23]. This motivates a careful consideration of fine-scale ancestry in the genetic analysis of complex Latino populations.

Investigations into the admixture of Mexicans have identified 2 primary subpopulations: nearly homogeneous Native Americans and a subset of ~80% of Mexican Mestizos [24, 25]. Mexican Mestizo individuals have varying ancestral proportions of Native American, European, and African ancestry, resulting in genetic heterogeneity both between and within Mestizos from different regions in Mexico [26]. Recently, it has been proposed that there are 7 admixture groups within Mexican Mestizos that are highly concentrated in specific geographic regions within Mexico, highlighting this population's diversity and genetic heterogeneity [27]. Two admixture groups, comprising >90% of the individuals, were identified in the southeast while, 2 other admixture groups were not represented in the north and central west [27]. Therefore, it is of vital importance to correct appropriately for population stratification in association studies performed

Color version available online

Fig. 1. NeuroMex study sites throughout Mexico will include Mexico City, Campeche, Querétaro, Jalisco, and Guanajuato, which are in distinct states in the country. NeuroMex, Neuropsychiatric Genetics Research of Psychosis in Mexican Populations.

in admixed populations, such as Mexican Mestizos, in particular when disease prevalence is affected by ancestry [28].

To more fully characterize the genetic architecture of psychosis in the Mexican population, the National Institute of Psychiatry Ramón de la Fuente Muñiz (INPRFM) in Mexico partnered with the Stanley Center for Psychiatric Research at the Broad Institute and the Harvard T.H. Chan School of Public Health (HSPH) to expand psychiatric genetics research in Mexico, launching the Neuropsychiatric Genetics Research of Psychosis in Mexican Populations (NeuroMex) project. The NeuroMex project aims to expand the knowledge of the genetic architecture of SCZ and BD in Mexico through large-scale sample collection, analysis, and follow-up, to increase the understanding of the genetics of Mexican populations, as well as to enhance neuropsychiatric genetic research capacity in Mexico through the training of scientists and support the development of locally led research programs.

Materials and Methods

Study Design Synopsis

NeuroMex is a case-control study, with cases defined as individuals with SCZ, schizoaffective, and BD (grouped under the heading "psychotic disorders") and controls defined as age and sex-matched individuals from the same geographic location without a history of psychosis.

Study Sites

The NeuroMex study is being conducted over 4 years, starting in April 2018 and ending in 2022, within 5 states throughout Mexico, including Mexico City, Campeche, Guanajuato, Jalisco, and Querétaro (Fig. 1). DNA extraction will be performed in the Pharmacogenetics Department at the INPRFM in México City, and genetic analysis will be performed at the Broad Institute in Cambridge, MA, USA.

Eligibility of the Study Sites Criteria

Sites were selected based on established working relationships between the INPRFM and the main public psychiatric hospital in the state. Each local collaborator identified a nearby general health facility willing to facilitate the collection of control samples, ensur-

ing that samples for cases and controls will be collected from the same underlying population and ancestry. In order to be considered as a local collaborator, sites needed to have adequate and available clinical and research personnel, clinical experience in psychotic disorders, and necessary research infrastructure, and draw patients representative of the local community. Finally, sites needed to demonstrate interest in developing capacities to further research on psychiatric genetics.

In Mexico City, the cases are recruited at the INPRFM, and controls at the Dr. Manuel Gea González General Hospital. The site in Campeche will recruit cases from the Psychiatric Hospital of Campeche and controls at the Dr. Javier Buenfil Osorio General Hospital of Specialties. In Querétaro, cases will be recruited at the State Center for Mental Health (CESAM) and controls at the Querétaro State Health Centers. In León, cases will be recruited at the Integral Health Care Center for Mental Health (CAISAME) and controls from the General Hospital of León, Guanajuato. In Jalisco, cases will be recruited from the CAISAME in Zapopan, Jalisco, and controls from Health Centers of the State of Jalisco, the Health Center of Zapote del Valle, and the Health Centers of Hacienda Santa Fe and Santa Cruz del Valle.

Sample

NeuroMex aims to enroll 9,208 participants, including 4,604 cases and 4,604 controls.

Cases

Inclusion criteria include (1) men or women, who have been clinically assessed and meet the DSM-IV and DSM-5 diagnostic criteria for SCZ, SAD, or BD type I ("psychotic disorders"); (2) age ≥18 years; (3) age of onset of the disorder <60 years; (4) Mexican patients, with Mexican parents and grandparents; and (5) voluntarily agrees to participate in the study.

Exclusion criteria include (1) institutionalized psychiatric patients; (2) experiencing acute levels of alcohol or under the influence of an illegal substance demonstrated by inpatient medical care for alcohol or substance abuse; (3) is currently experiencing involuntary detention; (4) participant is not fluent in the Spanish language which the consent form and phenotypic questionnaires are administered in; and (5) not competent to consent to the study, as defined by the UBACC [29].

Controls

Controls will consist of volunteers and individuals who present themselves or accompany others for treatment of general medical conditions at general hospitals that draw from similar catchment areas to the psychiatric hospitals where cases are recruited. Controls will be matched at the analysis stage to cases in accordance with age (within a 10-year age-group), sex, and recruitment site.

Inclusion criteria include (1) men or women with an age ≥18 years; (2) Mexican subjects, with Mexican parents and grandparents; and (3) voluntarily agrees to participate in the study. Exclusion criteria include (1) self-reported psychotic disorder; (2) experiencing acute levels of alcohol or under the influence of an illegal substance demonstrated by inpatient medical care for alcohol or substance abuse; (3) is currently experiencing involuntary detention; (4) participant is not fluent in the Spanish language which the consent form and phenotypic questionnaires are administered in; and (5) not competent to consent to the study, as defined by the UBACC [29].

Assessment for Cases and Controls

- Demographics: a number of baseline demographic variables will be collected including sex, age, personal and family place of birth, occupational status, educational attainment, civil status and living arrangements, native language and parent's native language, and socioeconomic status.
- MINI International Neuropsychiatric Interview (MINI): Standard 7.0.2 Modules A, C, K, and O on major depressive episode, manic and hypomanic episodes, and psychotic disorders and mood disorder with psychotic features, respectively [30].
- University of California, San Diego Brief Assessment of Capacity to Consent (UBACC), 10-item scale to assess decision-making capacity and evaluates a potential participant's understanding and appreciation of the NeuroMex study following a review of the study's consent form [29].
- The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST), a subscale developed to detect alcohol and substance use in primary and general medical care settings [31].
- Composite International Diagnostic Interview (CIDI), a checklist to identify chronic physical conditions such as diabetes, epilepsy, and heart disease [32].
- The standard Self Report version of the Life Events Checklist (LEC-5) scale used to screen for potentially traumatic events in a respondent's lifetime [33].
- Vital signs: blood pressure, heart rate, weight, height, BMI, and waist and hip diameter.
- Chart review and clinical diagnosis: participants will be asked about the history of their disease, and the information will be complemented with the clinical record. This includes current diagnosis, age of onset, age at first psychiatric treatment, initial polarity (BD patients only) suicidal ideation and attempts throughout life, relatives' mental and medical conditions, number of psychiatric hospitalizations throughout life, duration of symptoms, comorbid mental and medical diagnoses, and current and historic medication.

Consent Procedures

To ensure that participants (cases and controls) have the ability to consent and understand their participation in the study, the UBACC scale will be administered at maximum twice. Information previously misunderstood will be re-explained following the first administration and evaluated again during a second administration. The process ends once the full score of 20 is obtained or after the second trial, if the person has achieved a score of 14.5 or higher. If the person is unable to obtain a score of 14.5 following the 2 attempt, he/she will not be included within the study. This study has been approved by the Institutional Review Board at the INPRFM, the HSPH, and at each of the local institutions.

Procedures for Recruitment of Participants

The study flow is illustrated in Figure 2. Participants will be identified by psychiatric staff through a review of medical records and consultation with a clinician regarding their eligibility for the study. Control group participants will be recruited from general hospitals and general health facilities; study participation will take place on the clinic premises. A research assistant will consult the prospective participant in a private room at the clinic and carefully explain in Spanish the purpose and procedures of the study and emphasize that participation is entirely voluntary and will not impact any medical care they receive. Following a successful ad-

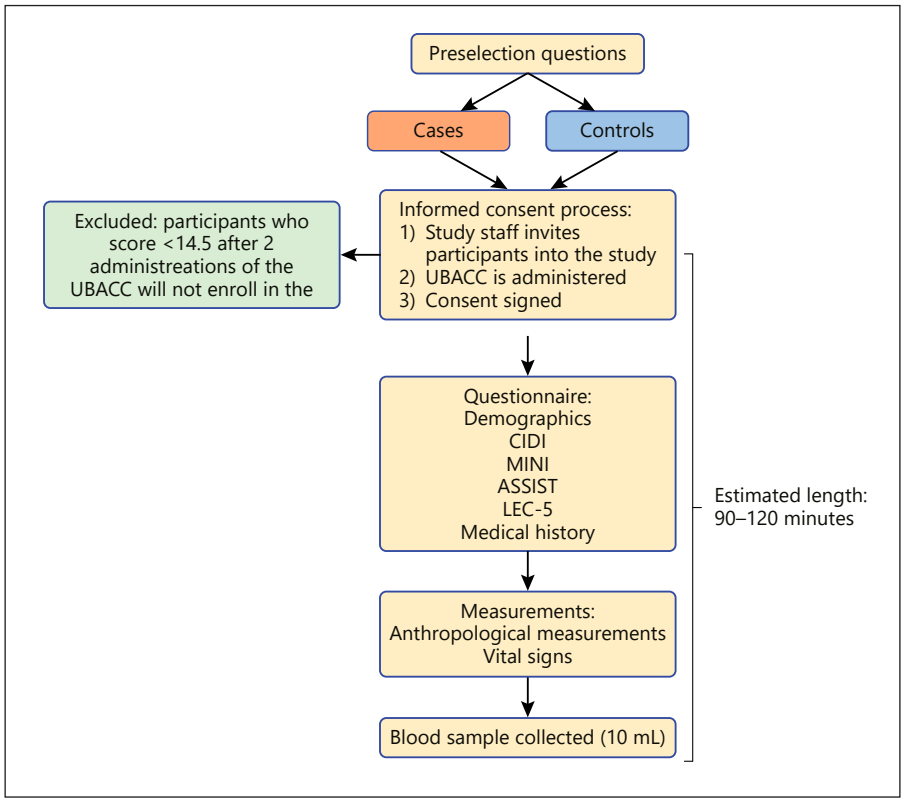


Fig. 2. Schematic illustrating the recruitment procedure for cases and controls in the NeuroMex study. This visualizes the informed consent process and the assessment of the clinical scales that will apply for the definition of a psychosis phenotype. UBACC, University of California, San Diego Brief Assessment of Capacity to Consent; CIDI, Composite International Diagnostic Interview Screener; MINI, Mini International Neuropsychiatric Interview, Standard 7.0.2 for Diagnostic and Statistical Manual of Mental Disorders-5; ASSIST, Alcohol, Smoking and Substance Involvement Screening Test, V.3.0; LEC-5, Life Events Checklist for Diagnostic and Statistical Manual of Mental Disorders-5; NeuroMex, Neuropsychiatric Genetics Research of Psychosis in Mexican Populations.

ministration of the UBACC and signed consent form, the research assistant will administer the questionnaires to collect phenotypic data.

All participants (cases and controls) will be assessed with the MINI to measure positive and negative symptoms as well as mania. Additionally, all participants will receive the Chronic Conditions Screener from CIDI, an excerpt from the ASSIST relating to alcohol and substance use, LEC-5, and have their vital signs taken, including blood pressure, heart rate, height, and weight. Phenotypic data will be collected by a research assistant on encrypted tablets using secure software designed for the collection of Research Electronic Data Capture (REDCap) [34]. Phenotypic data will be collected from patient charts and from patient interviews following consent. Finally, a sample of 10 mL of blood will be drawn in a private room by a laboratory technician. Whenever possible, the research visit will be scheduled to coincide with any clinical visit requiring a blood draw, to reduce the number of pricks.

Control study participants will be compensated upon completion of the study assessments in Mexico City and Campeche via 2 movie ticket vouchers, valued at 50 Mexican pesos each. In Querétaro, León, and Zapopan, none of the study participants will receive compensation for participation in the study in accordance with local research norms.

DNA Extraction and Processing

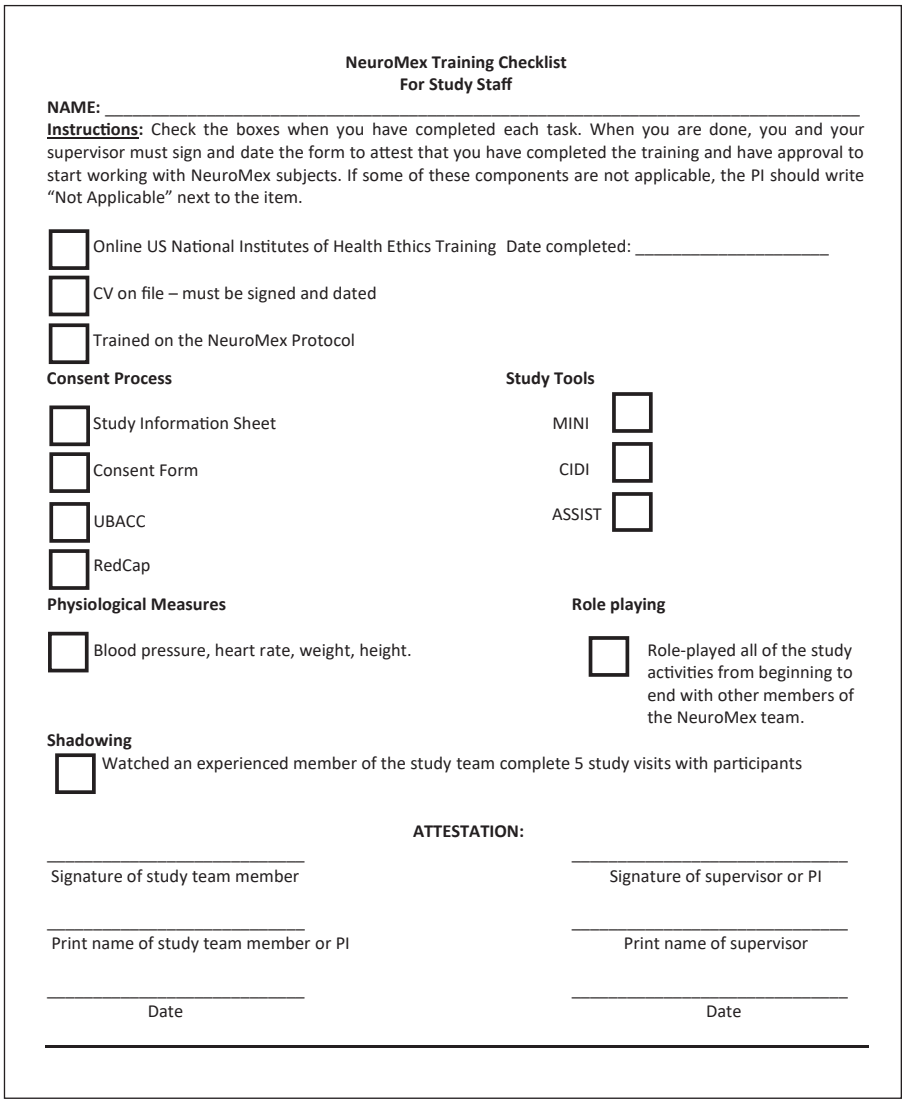
All participants provide 10 mL blood sample in vacutainer tubes with EDTA which is shipped to the Clinical Laboratory at the INPRFM in Mexico City, Mexico. Genomic DNA is isolated from EDTA blood samples using the FlexiGene DNA kit (QIAGEN, Cat.51206). The DNA obtained is quantified and adjusted in a concentration of 500 ng/μL. An aliquot of the extract in a concentration of 100 ng/μL is sent to the Broad Institute of MIT and Harvard in Cambridge, MA, USA. The remaining DNA will be frozen and stored in the Pharmacogenetics Department at INPRFM in Mexico City within a designated biobank and will be managed and kept in accordance with the rules of the institution.

Training and Capacity Building

The research study staff involved in the recruitment of participants will be certified by the online training program Collaborative Institutional Training Initiative and the online Human Subject Training offered by US National Institutes of Health. Study staff will also undergo multiple training sessions to ensure their understanding of the study and their role. Additionally, they will undergo extensive role-playing and clinical training in the application of the MINI, ASSIST, LEC-5, and UBACC scales.

Harvard staff will provide in-person training to the clinical coordinators at the INPRFM, providing an overview of the project, role-playing potential interactions with participants related to ethics and confidentiality. Additionally, clinical coordinators will be trained how to properly recruit and consent study participants, report adverse events, and properly store study forms and data in accordance with procedures established by the HSPH Quality Improvement Program (QIP).

In order to ensure adequate study staff training, a training checklist must be completed before research staff is allowed to recruit study participants (Fig. 3). This form will require the signature and date of each study staff supervisor before the staff member



NeuroMex Training Checklist
For Study Staff

NAME: _____

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 NeuroMex 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 NeuroMex Protocol

Consent Process

☐ Study Information Sheet

☐ Consent Form

☐ UBACC

☐ RedCap

Study Tools

MINI ☐

CIDI ☐

ASSIST ☐

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 NeuroMex team.

Shadowing

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

ATTESTATION:

_____ Signature of study team member

_____ Print name of study team member or PI

_____ Date

_____ Signature of supervisor or PI

_____ Print name of supervisor

_____ Date

Fig. 3. NeuroMex training checklist. ASSIST, Alcohol, Smoking and Substance Involvement Screening Test, V.3.0; CIDI, Composite International Diagnostic Interview screener; LEC-5, Life Events Checklist for Diagnostic and Statistical Manual of Mental Disorders-5; MINI, Mini International Neuropsychiatric Interview, Standard 7.0.2 for Diagnostic and Statistical Manual of Mental Disorders-5; PI, Principal Investigator; UBACC, University of California, San Diego Brief Assessment of Capacity to Consent; NeuroMex, Neuropsychiatric Genetics Research of Psychosis in Mexican Populations.

begins recruiting participants. All study staff certifications will be uploaded to the study's regulatory binder based on QIP's recommendations and instructions.

The research and laboratory staff will be trained to enhance neuropsychiatric genetic research capacity in Mexico and support the development of locally led research programs. All collaborators will receive training on the use, access, and analysis of study data obtained during the NeuroMex study and also by taking part in workshops and fellowships offered through the training and capacity building program, the Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER).

Data Management and Security

Phenotypic data will be collected on tablets using the REDCap application hosted by the HSPH. New participant data, all de-identified, will be uploaded to HSPH's secure server at the end of each day. All identifiable data will stay at each local institution in Mexico.

Once per week, data will be downloaded from REDCap to a Google Cloud platform (GCP) MySQL 5.7 instance for quality control (QC) and efficient storage. The data will not be removed from REDCap, and any changes will be captured by the application's built-in audit log. When corrections are made (e.g., a patient is recontacted because a question was skipped), they are tracked by study staff in separate logs independent of REDCap. When an update is made in REDCap, the outdated observation will first be removed from GCP storage. Corrected records will be included with new ones during the weekly download and undergo QC again. The REDCap Application Programming Interface will be leveraged in Python 3 to automate and standardize these processes.

Data will be distributed on the Broad Institute's Terra platform. Terra is built on GCP's secure infrastructure and implements further security measures managed by the Broad Institute's Information Technology staff. It is possible to distribute both genetic and phenotype data through Terra; users can access either a full or partial dataset, but require explicit approval.

Genetic Data Analysis Plan: Processing and QC of Pilot Genetic Data

QC procedures for NeuroMex data will use the Hail Python library [35] following gold standard procedures. All data will be stored in the cloud on the Google Cloud and Terra platforms. The following QC steps and filters have been adapted into a custom pipeline for use on NeuroMex data, based on guidelines set by previous efforts such as RICOPOLI [36, 37]. This pipeline is fully scalable and will be made publicly available on github to support other analyses of diverse Latina populations. We will additionally provide a Wiki/documentation alongside the code to ensure all users will be able to utilize it fully.

The following list of QC steps and parameters will be used to assess NeuroMex genetic data: (1) remove variants with a call rate <95%; (2) remove individuals with a call rate <98%; (3) remove individuals with an inbreeding coefficient above 0.2 and below −0.2; (4) remove individuals whose reported sex did not match their genotypic sex; (5) remove variants with a call rate <98% after individual-level filtering; (6) remove variants with a minor allele frequency <0.5%; and (7) remove variants with a Hardy-Weinberg equilibrium p value $<1 \times 10^{-3}$.

Phasing and Imputation

The data will be phased using EAGLE2 [38] prior to imputation with minimac3 [39], using the data jointly called with other relevant reference panels, where feasible [40], such as the Human Genome Reference Panel (HGDP) [41]. Together, this jointly called reference panel will contain a high representation of diverse ancestry individuals from across the Americas, relevant for these cohorts. A joint-called dataset of the HGDP + 1000 G is in process at the Broad Institute for these purposes.

GWAS for Psychosis

The primary goal of our proposal is to find loci that influence susceptibility to SCZ in Mexican populations. We will first conduct a GWAS to discover variants associated with risk of psychosis in the NeuroMex dataset. We will conduct GWAS in 2 ways. First, we will run a traditional logistic regression analysis using 10 principle components to correct for population structure. This will provide aggregate summary statistics for the Mexican population and provide a point of comparison to what is traditionally done in the field. Second, we will utilize the new Tractor method [42] to generate ancestry-specific summary statistics conditioned on the local ancestry background within admixed individuals. This will involve conducting local ancestry inference using RFmix [43], followed by a local ancestry informed logistic regression model. Tractor runs will allow us to identify novel hits driven by specific ancestry components within our admixed cohorts. The primary GWAS discovery analyses will be conducted for the dichotomous psychosis diagnosis using a generalized linear mixed model implemented in SAIGE [44]. We will include age, sex, and genetic ancestry covariates estimated with principal components analysis (PCA). All analyses will be performed using Hail [43], PLINK [45], python, and R. To assess significance thresholds and correct for multiple comparisons, we will apply the conventional genome-wide significance threshold of $p < 5 \times 10^{-8}$.

Following our GWAS, we will combine our summary statistics with that of the Psychiatric Genetics Consortium SCZ and BD working groups using inverse-variance weighted meta-analysis to increase our power to discover variants shared across populations. Last, we will fine-map associations that influence risk. Multi-ancestry methods for fine-mapping will help resolve our association signals to pinpoint putative variants or smaller regions affecting an individuals' risk of psychosis.

Estimation of Heritability

We will estimate the fraction of disease risk attributable to inherited genetic differences via SNP-based heritability estimates. These estimates will inform the power of the GWAS. We will estimate heritability using genome-wide complex trait analysis [46] and an extension of LD score regression (cov-LDSC) designed for admixed populations [47] to estimate SNP heritability within Mexico. We will assess genetic correlation with existing GWAS in other globally diverse populations using POPCORN [48], which will help inform both the genetic correlations as well as the phenotypic heterogeneity across populations.

Polygenic Risk Scoring

Given that polygenic risk scores (PRSs) do not transfer well across ancestries [49], we will assess the predictive accuracy of PRS built from the largest GWAS of psychosis – consisting of primarily European individuals – on our cohort. We will additionally train and test a model using our own cohort data to compare predictive performance on a much smaller but better matched sample. For this, we will create a hold-out testing cohort comprising 10% of our sample. We will use the remaining 90% of our sample in the generation of summary statistics on which we will build a PRS model in PLINK. We will then assess prediction accuracy in our testing cohort as compared to prediction accuracy using traditional PRSs built off of European summary statistics. Additionally, we may assess the predictive accuracy of a novel method designed for calibrating European-derived summary statistics for use in admixed cohorts [50]. This method scales prediction based on the local ancestry composition of individuals in the cohort and therefore makes scores transfer more appropriately across ancestry groups.

Population Structure and Admixture

We will examine population composition in the NeuroMex samples using allele frequency-based approaches, including ADMIXTURE and PCA [51, 52]. We will assess the best fit number of ancestry clusters in our cohort with ADMIXTURE using 5-fold cross-validation. For PCA analyses, we will project samples onto PC space as defined by our 1,000 Genomes and HGDP reference panel data to capture global diversity. We will also investigate subcontinental diversity by comparing genetic diversity with relevant reference populations from the Americas and ancestral origins. To measure ancestry-specific subcontinental diversity, we will run ancestry-specific PCA, for example, on Native American, European, and African haplotypes [53]. We also plan to investigate the local ancestry composition of individuals across Mexico using RFmix [43]. This will shed light on the population history of these groups and inform the appropriate parameters to use in the ultimate association analysis.

Discussion

The primary aim of the study was to contribute to the identification of genetic variants associated with psychosis in the Mexican population through large-scale sample

collection and analysis. GWASs have reported primarily on European ancestry populations; therefore, patients of non-European ancestry are under-represented in the findings regarding the genetic architecture of psychotic disorders. Via the inclusion of diverse ancestry participants, with the purpose to identify shared and novel variants affecting psychosis, this study will begin to fill in the missing heritability that may be explained by under-representation of non-European populations within GWAS cohorts [22, 54]. The exclusion of non-European populations limits the extensibility of GWAS findings and underestimates the genetic burden carried by individuals in diverse populations [55]. Other groups also have unique genetic variation that can better inform the genetic basis of psychosis and aid in fine-mapping association signal. For example, a GWAS identified specific population risk variants associated with SCZ in a Chinese population as well as loci previously found in GWAS including only individuals of European ancestry [56].

There is a significant genetic variation worldwide; therefore, findings cannot be extrapolated a priori across populations. For example, there are differences in the allele distributions of triallelic 5-HTTLPR/rs25531 variants across continental groups. Murphy et al. [57] showed the frequency of low-activity alleles (S/L_G) in Europeans is 22%, in American-Indians 43%, and in Asians 60%. It was showed that Mexican Mestizos have a frequency of 57% [58]. Therefore, the inclusion of several ancestries can identify variants that would not have been detected in studies composed solely of samples from a population of European ancestry [59, 60]. As such, this study could help identify new loci associated with psychosis in an under-represented population which will improve the understanding of this disease in Mexican populations and expand the knowledge regarding the genetics of psychosis for all ancestry groups.

It is unknown if Latino American psychotic patients share the same identified genetic risk factors in European population. Additionally, although the impact of an adverse environment in the development of the psychiatric disorders is unclear, it could influence differences between Latino American and Latin American psychotic patients. Latin Americans report the highest overall rates of trauma exposure, poverty, and limited social and health-care assistance [61]. The NeuroMex study will include the LEC-5 scale to screen for potentially traumatic events with the intent to analyze the impact of environmental risk factors on the psychotic disorders. The present study could help understand the differences and similarities between Latino Americans and European populations and increase the ethnic diversity of samples used when studying the genetics of psychotic disorders.

The NeuroMex study will be the first large-scale GWAS to address the genetic architecture of SCZ and BD ("psychotic disorders") in the Mexican population. Furthermore, it will be the first psychiatric genetics study that incorporates collaboration with psychiatric hospitals of the Mexican Republic with the purpose of forming a national research consortium. Finally, it will train researchers locally who will conduct their own research in the future.

The study has many strengths, including that the clinical staff are psychiatrists with expertise in SCZ and BD. Additionally, the clinical staff will be trained to ensure an adequate and unified diagnosis among all sites. However, there are also several limitations that need to be considered. First, the small sample size will result in limited statistical power; and second, samples are not comprehensively representative of the Mexican population. The sampling of participants from additional regions of Mexico would represent a relevant research study that would ensure that findings contribute to the characterization of the genetic factors underlying the psychotic disorders in the Mexican population.

Recently, it was observed that Mexican Mestizos is a heterogeneous group composed of 7 admixture groups distributed in 5 geographical regions of Mexico [27]. Our sample will be represented for 3 regions, in the central west with Jalisco; central east with Mexico City and Querétaro; and southeast with Campeche, each 1 covering different admixture patterns. However, in the north and northwest, 2 admixture groups were identified as most closely related to the Native American ancestry [27, 62]; regions that are not currently included in the NeuroMex study.

A second potential limitation is the clinical heterogeneity of SCZ and BD. It has been suggested that the definition of particular endophenotypes that cut across psychiatric disorders' diagnoses can help create clinically meaningful phenotypes that have the potential to identify genetic discovery.

This study will provide valuable information regarding the genetic architecture of SCZ and BD in the Mexican population. The findings have the potential to provide evidence of genetic markers shared across populations and lead to the identification of loci restricted to Mexican population.

Statement of Ethics

The study protocol was developed in accordance with the Declaration of Helsinki and has been approved by the Ethics Committees from all participating sites: Mexico City: Research Ethics Committee from the INPRFM (09-CEI-010-20170316, approval number: CEI/017/2016); Campeche: Campeche State Bioethics Commission; Queretaro: the Ethics Committee from Secretary of Health of Querétaro (approval number: 924/CESAM/21-04-2017); León: the Ethics Committee (CONBIOETICA-11-CEI-003-20190704); Jalisco: the Ethics Committee of Jalisco (CEI/30032021); and the USA: the HSPH (approval number: IRB17-1914). Written informed consent will be obtained from all participants. The research study staff will use the UBACC scale during the consent process to make sure the participants understand the study, the procedures that involve their participation, possible side effects, and that they can withdraw at any point. The confidentiality of participants will be instituted using 3-ID part system in order to prevent the possibility of connecting the participant with their data. One ID for phenotypic information, 1 ID for contact information, and another ID for genetic information will be assigned with a randomly generated numerical alpha code that will be identify with a QR code. A single encrypted database linking the 3 IDs will be stored in-country. The consent informed form and MINI scale will be stored in a locked cabinet.

Conflict of Interest Statement

A.R.M. serves as a consultant for 23andMe and is a member of the Precisely Scientific Advisory Board. The remaining authors declare no competing interests. The authors declare that they have no competing interests.

Funding Sources

This study is funded by the Stanley Center for Psychiatric Research at Broad Institute. E.G.A. and A.R.M. were supported by funding from the National Institutes of Health (K01MH121659, T32MH017119; and K99MH117229, respectively). The sponsor was involved in the study design, in the writing of the report, and in the decision to submit the manuscript for publication. The sponsor will be involved in the analysis and interpretation of the data.

Author Contributions

All the authors contributed to the conception and design of this manuscript. K.S.K. and B.C. design the study. B.C., B.F.S., E.G.A., A.R.M., and A.M.O. wrote the first draft of the manuscript. P.M.C., J.J.P., C.B.P., R.E., M.B., L.B.C., H.O., A.M.O., Z.K., R.E.B., and C.M.L., contributed in the process of drafting and revising. All the authors gave their agreement and approval for all aspects of the final version of the manuscript.

Data Availability Statement

Genetic information will be available to third-party researchers according to data sharing policies to be determined by study investigators.

References

1 van Os J, Hanssen M, Bijl RV, Ravelli A. Strauss (1969) revisited: a psychosis continuum in the general population? Schizophr Res. 2000 Sep;45(1–2):11–20.
2 Hill SK, Reilly JL, Keefe RS, Gold JM, Bishop JR, Gershon ES, et al. Neuropsychological impairments in schizophrenia and psychotic bipolar disorder: findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. Am J Psychiat. 2013 Nov;170(11):1275–84.
3 Cardno AG, Rijsdijk FV, Sham PC, Murray RM, McGuffin P. A twin study of genetic relationships between psychotic symptoms. Am J Psychiat. 2002 Apr;159(4):539–45.
4 van Os J, van der Steen Y, Islam MA, Gülöksüz S, Rutten BP, Simons CJ, et al. Evidence that polygenic risk for psychotic disorder is expressed in the domain of neurodevelopment, emotion regulation and attribution of salience. Psychol Med. 2017 Oct;47(14):2421–37.
5 Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry. 2015 Apr;72(4):305–15.
6 Shevlin M, McElroy E, Bentall RP, Reininghaus U, Murphy J. The psychosis continuum: testing a bifactor model of psychosis in a general population sample. Schizophr Bull. 2017 Jan;43(1):133–41.
7 Guloksuz S, van Os J. The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. Psychol Med. 2018 Jan;48(2):229–44.
8 Cardno AG, Owen MJ. Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. Schizophr Bull. 2014 May;40(3):504–15.
9 Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 2009 Jan;373(9659):234–9.
10 Van Snellenberg JX, de Candia T. Meta-analytic evidence for familial coaggregation of schizophrenia and bipolar disorder. Arch Gen Psychiatry. 2009 Jul;66:748–55.
11 International Schizophrenia Consortium; Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug;460(7256):748–52.
12 Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell. 2019 Dec;179(7):1469–82.e11.
13 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014 Jul;511(7510):421–7.
14 Bigdeli TB, Genovese G, Georgakopoulos P, Meyers JL, Peterson RE, Iyegbe CO, et al. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. Mol Psychiatry. 2020 Oct;25(10):2455–67.

15 Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019 May; 51(5):793–803.
16 Zhuo C, Hou W, Li G, Mao F, Li S, Lin X, et al. The genomics of schizophrenia: shortcomings and solutions. Prog Neuropsychopharmacol Biol Psychiatry. 2019 Jul;93:71–6.
17 Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM. Predicting polygenic risk of psychiatric disorders. Biol Psychiatry. 2019 Jul;86(2):97–109.
18 Escamilla MA, Ontiveros A, Nicolini H, Raventos H, Mendoza R, Medina R, et al. A genome-wide scan for schizophrenia and psychosis susceptibility loci in families of Mexican and Central American ancestry. Am J Med Genet B Neuropsychiatr Genet. 2007 Mar;144B(2):193–9.
19 Gonzalez S, Camarillo C, Rodriguez M, Ramirez M, Zavala J, Armas R, et al. A genome-wide linkage scan of bipolar disorder in Latino families identifies susceptibility loci at 8q24 and 14q32. Am J Med Genet B Neuropsychiatr Genet. 2014 Sep;165B(6):479–91.
20 Gonzalez S, Gupta J, Villa E, Mallawaarachchi I, Rodriguez M, Ramirez M, et al. Replication of genome-wide association study (GWAS) susceptibility loci in a Latino bipolar disorder cohort. Bipolar Disord. 2016 Sep;18(6):520–7.
21 Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019 Jan;47(D1):D1005–12.
22 Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016 Oct;538(7624): 161–4.
23 Spear ML, Diaz-Papkovich A, Ziv E, Yracheta JM, Gravel S, Torgerson DG, et al. Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. Elife. 2020 Dec;9:e56029.
24 Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, et al. Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet. 2008 Mar;4(3):e1000037.
25 Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, et al. Human genetics. The genetics of Mexico recapitulates native American substructure and affects biomedical traits. Science. 2014 Jun;344(6189):1280–5.
26 Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, Fernandez-Lopez JC, Uribe-Figueroa L, Contreras A, et al. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. Proc Natl Acad Sci U S A. 2009 May;106(21):8611–6.
27 Martínez-Magaña JJ, Genis-Mendoza AD, Villatoro Velázquez JA, Camarena B, Martín Del Campo Sanchez R, Fleiz Bautista C, et al. The identification of admixture patterns could refine pharmacogenetic counseling: analysis of a population-based sample in Mexico. Front Pharmacol. 2020 Apr;11:324.
28 Huerta-Chagoya A, Moreno-Macías H, Fernández-López JC, Ordóñez-Sánchez ML, Rodríguez-Guillén R, Contreras A, et al. A panel of 32 AIMs suitable for population stratification correction and global ancestry estimation in Mexican mestizos. BMC Genet. 2019 Jan;20(1):5.
29 Jeste DV, Palmer BW, Appelbaum PS, Golshan S, Glorioso D, Dunn LB, et al. A new brief instrument for assessing decisional capacity for clinical research. Arch Gen Psychiatry. 2007 Aug;64(8):966–74.
30 Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, 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 ICD10. J Clin Psychiatry. 1998;59(Suppl 20):22–33.
31 Humeniuk R, Ali R, Babor TF, Farrell M, Formigoni ML, Jittiwutikarn J, et al. The alcohol, smoking and substance involvement screening test (ASSIST). Addiction. 2008 Jun; 103(6):1039–47.
32 Kessler RC, Ustün TB. The world mental health (WMH) survey initiative version of the world health organization (WHO) composite international diagnostic interview (CIDI). Int J Methods Psychiatr Res. 2004;13(2):93–121.
33 Weathers FW, Blake DD, Schnurr PP, Kaloupek DG, Marx BP, Keane TM. The Life Events Checklist for DSM-5 (LEC-5). Instrument available from the National Center for PTSD. 2013. www.ptsd.va.gov.
34 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009 Apr;42(2):377–81.
35 Hail. n.d. Github. https://github.com/hail-is/hail Accessed 2020 Jan 30.
36 Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. RICOPILI: rapid imputation for consortias pipeline. Bioinformatics. 2020 Feb;36(3): 930–3.
37 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9): 1564–73.
38 Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef A, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016 Nov; 48(11):1443–8.
39 Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Aug;48(10):1284–7.
40 The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 Oct;526:68–74.
41 Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020 Mar;367(6484):eaay5012.
42 Atkinson EG, Maihofer AX, Kanai M, Martin AR, Karczewski KJ, Santoro ML, et al. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. Nat Genet. 2021 Feb;53(2):195–204.
43 Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013 Aug;93(2): 278–88.
44 Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nature Genet. 2018 Sep; 50(9):1335–41.
45 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.
46 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan;88(1): 76–82.
47 Luo Y, Li X, Wang X, Gazal S, Mercader JM; 23 and Me Research Team; SIGMA Type 2 Diabetes Consortium, et al. Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. Hum Mol Genet. 2021 Jul 28;30(16):1521–34.
48 Brown BC; Asian Genetic Epidemiology Network Type 2 Diabetes Consortium; Ye CJ, Price AL, Zaitlen N. Transethnic genetic-correlation estimates from summary statistics. Am J Hum Genet. 2016 Jul;99(1):76–88.
49 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51:584–91.
50 Marnetto D, Pärna K, Läll K, Molinaro L, Montinaro F, Haller T, et al. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. Nat Commun. 2020 Apr 2;11(1):1628.
51 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.

52 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep; 19(9):1655–64.
53 Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al. Reconstructing the population genetic history of the Caribbean. PLoS Genet. 2013 Nov;9(11): e1003925.
54 Dennison CA, Legge SE, Pardiñas AF, Walters JTR. Genome-wide association studies in schizophrenia: recent advances, challenges and future perspective. Schizophr Res. 2020 Mar;217:4–12.
55 Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet. 2017 Apr;100(4):635–49.
56 Li Z, Chen J, Yu H, He L, Xu Y, Zhang D, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat Genet. 2017 Nov;49(11):1576–83.
57 Murphy DL, Maile MS, Vogt NM. 5HTTLPR: white knight or dark blight? ACS Chem Neurosci. 2013 Jan;4(1):13–5.
58 Camarena B, Hernández S, González L, Griselda F, David L, Aguilar A, et al. Association study between the triallelic polymorphism of the SLC6A4 gene and eating disorders. Am J Psychiatry Neurosci. 2018 Dec; 6(4):104–7.
59 Yu H, Yan H, Li J, Li Z, Zhang X, Ma Y, et al. Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. Mol Psychiat. 2017 Jul;22(7):954–60.
60 Ohi K, Shimada T, Yasuyama T, Uehara T, Kawasaki Y. Variability of 128 schizophrenia-associated gene variants across distinct ethnic populations. Transl Psychiatry. 2017 Jan;7(1): e988.
61 Fonseca L, Sena BF, Crossley N, Lopez-Jaramillo C, Koenen K, Freimer NB, et al. Diversity matters: opportunities in the study of the genetics of psychotic disorders in low- and middle-income countries in Latin America. Braz J Psychiatry. 2020 Nov; S1516-44462020005038202.
62 Costa-Urrutia P, Abud C, Franco-Trecu V, Colistro V, Rodríguez-Arellano ME, Alvarez-Fariña R, et al. Effect of 15 BMI-associated polymorphisms, reported for Europeans, across ethnicities and degrees of Amerindian ancestry in Mexican children. Int J Mol Sci. 2020 Jan;21(2):374.