Diversity matters: opportunities in the study of the genetics of psychotic disorders in low- and middle-income countries in Latin America

Lais Fonseca, Brena F. Sena, Nicolas Crossley, Carlos Lopez-Jaramillo, Karestan Koenen, Nelson B. Freimer, Rodrigo A. Bressan, Sintia I. Belangero, Marcos L. Santoro, Ary Gadelha

Introduction

Psychotic spectrum disorders – mainly schizophrenia, schizoaffective, and bipolar disorder – have an estimated lifetime prevalence of around 3.5% and represent a major global health burden, both at individual and public levels. Despite their social relevance, the pathophysiological pathways leading to psychotic disorders remain mostly uncertain. The development of research on the etiological features of psychosis is a crucial step towards more effective and improved diagnostic tools and treatments. To date, the most established hypothesis depicts psychotic disorders as a byproduct of gene-environment interactions during neurodevelopment. Thus, a broader view of genetic features and environmental exposures can provide new insights into the underlying neurobiology.

Most genetic studies on psychosis, however, have been conducted in high-income countries, using essentially European descendants. This lack of ethnic diversity stands out for large-scale genome-wide association studies (GWAS) in general, as more than 78% of these samples are from European ancestry. Consequently, other groups remain widely underrepresented. Latin Americans comprise merely 1.3% of all GWAS samples, despite accounting for 8.4% of the world population, which leads not only to serious ethical qualms, but also to critical scientific repercussions, since the same GWAS findings may not be replicable across ethnic groups.

This article highlights the importance of investigating the genetics of psychotic disorders in low- and middle-income countries (LMIC) from a Latin American perspective. We also discuss how this genetic background

Keywords: Latin America; psychotic disorders; schizophrenia; genetics; environment
interaction with the deprived setting of this region could benefit research. Finally, we review Latin America’s participation in genetic studies so far and describe three ongoing projects related to psychotic disorders in Latin America.

How genetic studies in Latin American populations can improve our understanding of psychotic disorders

Latin America encompasses 20 countries and over 650 million citizens in a vast geographic region, each with its own heterogeneous demographic features. From the admixture point of view, Latin America provides a highly diversified genetic pool, stemming from Native American, European, African, and Asian ancestry components that resulted from unprecedented transcontinental migrations since the 15th century.

The distribution of these ancestries, however, is extremely heterogeneous across the continent. Populations from Mexico, Peru, Ecuador, and Bolivia present over 40% Native American ancestry, whereas those of Uruguay, Argentina, and southern Brazil have less than 10%. Likewise, African ancestry is disproportionately encountered due to slavery and the colonial activities of European countries. Consequently, the African ancestry in Latin America ranges from Caribbean countries (mainly Haiti and Jamaica), which have major African contributions, to Argentina and Uruguay, for instance, whose populations have less than 5% African ancestry.

A more in-depth look into African migration illustrates the complexity of ethnic diversity across the region, since its distributions also vary considerably within each country: in Brazil, for example, the population of Salvador (Northeast region) presents 50% African ancestry, the largest in the country, whereas Southern Brazil has an African ancestry lower than 10%. These differences usually reflect the economic activities that motivated large, forced migrations. There are also ethnic differences within African ancestry, with greater West African ancestry in the Northeast of Brazil and East African ancestry in the Southeast region. In general, the same applies to European ancestry and migration flows from other continents.

One characteristic that usually distinguishes Latin Americans is their pronounced interethnic admixture. This contrasts with the genetic makeup of European populations, which limits the applicability of current genetic findings. Polygenic predictors of risk, namely the Polygenic Risk Score (PRS), represent a key instrument to approach complex disorders. PRS summarizes the genome-wide effects of thousands genetic variants for a specific trait and, in the last decade, have allowed substantial insights into complex illnesses. But these scores are mainly derived from European samples, and such reduced ethnic diversity weakens the translation of results to non-European populations. For instance, among admixed populations, particularly those enriched with African ancestry, PRS has shown prediction accuracies up to 75% lower than in Europeans.

Studying more diverse samples would enhance the applicability of research findings. But, ultimately, a primary reason for including other populations is to obtain a representation of different allelic variants and, above all, allele frequencies different from those encountered in European samples. In fact, the inclusion of non-European subjects has shown to increase the number of identified genomic associations and aid in the fine-mapping of GWAS loci. The few studies that did include non-European populations discovered variants associated with a wide range of phenotypes. The 1000 Genomes (1000G) Project Consortium found that, in general, individuals from African populations present 800,000 more variant sites than European and Asian populations. Likewise, studies have shown that admixed populations with African ancestry presented the greatest variability per individual regarding the number of variants (for example, Colombians and Puerto Ricans from 1000G).

Notwithstanding the clear advantages of increasing diversity, admixed subjects are consistently removed from GWAS, mainly because population substructure can bias results even after global ancestry or principal component adjustments. However, new tools and statistical methods developed in the last few years to approach ancestry more effectively, such as Tractor and asaMap, should encourage the inclusion of admixed populations in large genomic studies.

How the study of the Latin American environment can improve our understanding of psychotic disorders

While the genetic contribution to psychosis is well-established, it does not fully explain the overall risk for these conditions. Genetic risk can be a potential catalyst for the independent effects of pathogenic environmental exposures and vice versa.

Some environmental risk factors for psychotic disorders seem to be particularly frequent and intense in Latin America, i.e., violence, poverty, and urbanicity. According to the United Nations International Children’s Emergency Fund (UNICEF), the overall wealth inequality in Latin American countries is the highest in the world and is directly linked to increased poverty, unregulated urbanization, crime, violence, and limited social or healthcare assistance. A recent report showed that, outside war zones, 43 of the world’s 50 most dangerous cities are located in Latin America, with one-third of the world’s homicides in 2018 occurring in this region. Latin America also has the world’s highest rates of interpersonal violence and kidnapping. Consequently, social inequality provides a natural experiment in which impoverished communities that are highly exposed to violence coexist within the same...
scores, which, analogous to the PRS (although more is through recently described poly-environmental risk possible approach to handling this myriad of effects opportunities to study gene-environment interaction. One differently by Latin America’s adverse environment disease severity, or age of onset could be impacted to trauma exposure.\textsuperscript{39} suggests that Latin American populations may be desensitized and/or environmentally acquired mechanisms, which sug-\textsubscript{39} result of cross-population variability in specific psychometric similarities or even higher than that in LMIC, which could be a traumatic stress disorder in high-income countries was highest overall rates. Unexpectedly, the prevalence of post-traumatic stress disorder in high-income countries was similar or even higher than that in LMIC, which could be a result of cross-population variability in specific psychometric instruments. On the other hand, these findings could also reflect resilience traits, with possible genetic associations and/or environmentally acquired mechanisms, which sug-\textsubscript{39} that Latin American populations may be desensitized to trauma exposure.

However, it is unclear how psychosis incidence rates, disease severity, or age of onset could be impacted differently by Latin America’s adverse environment compared to high-income countries, indicating further opportunities to study gene-environment interaction. One possible approach to handling this myriad of effects is through recently described poly-environmental risk scores, \textsuperscript{44,45} which, analogous to the PRS (although more dynamic), could load multiple environmental risk factors and express the impact of cumulative exposure, which is also in line with the syndemic framework. Additionally, the relative weights of each risk factor can be calculated for different regions and aggregate scores (as opposed to individual risks) and can be used to compare the overall environmental risk for psychosis. Like the previously discussed genetic findings, investigating environmental exposures in different regions could reveal new, previously undescribed risk factors and contribute to the development of more accurate, reproducible measures.

\textbf{Previous Latin American research}

Before genome-wide investigations, candidate gene studies with a case-control design were the main approach to exploring associations between single genomic loci and psychotic disorders. Even though the results have been, in general, mostly non-replicable,\textsuperscript{46} innumerous genes and variants have been tested, resulting in the identification of some small effect alleles.

Recently, Liu et al.\textsuperscript{47} conducted an extensive systematic review of schizophrenia candidate gene studies, which investigated a total of 20,570 single nucleotide polymorphisms (SNPs) in more than 3,414 genes. The authors made all the results available, which comprised the raw data of every study, including the country of origin and ethnicity of each sample. Among the 1,188 eligible publications, only 19 were from Latin America, which represents 1.5% of the included reports: Brazil (13 studies\textsuperscript{48-69}), Mexico (four studies\textsuperscript{60-63}), and Costa Rica (two studies\textsuperscript{64,65}), although the samples overlapped considerably. Overall, the 19 studies assessed seven cohorts, all with relatively small sample sizes, totaling 1,426 cases (range = 69-421; standard deviation [SD] = 120.1) and 2,020 controls (range = 85-834; SD = 298.0), whose ethnicities were described as “Mixed.”

Other Latin American studies have gathered samples larger than 1,000 subjects and performed candidate loci/ gene associations. Gonzalez et al.\textsuperscript{66} evaluated 91 SNPs previously associated with schizophrenia or bipolar disorder in a Latin American cohort of 2,254 individuals (706 cases of bipolar disorder or schizoaffective disorder and 1,548 probands) and were able to replicate the association of eight SNPs. We found that 80% of these studies aimed at replicating the findings of other populations, rather than investigating new genetic associations.

To date, Bigdeli et al.\textsuperscript{22} have conducted the only Latin American GWAS with reasonable power, which included 4,324 individuals (1,234 cases). This study identified a novel genome-wide significant association within the \textit{GALNT13} gene in the Latino cohort alone and 101 loci when meta-analyzing it with the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC-Schizophrenia),\textsuperscript{4} eight of which had not been previously associated with schizophrenia. Moreover, Bigdeli et al. showed that meta-analyzing combined data from European (PGC-Schizophrenia) admixed African and Latinos improved the fine-mapping of nine regions and increased PRS prediction for Latinos from 1.7 to 2.1% (measured by liability \textit{R}²).

Recent technologies have allowed the scanning of the entire exome, allowing the identification of rare genetic variants, which could lead to larger effects in deleterious phenotypes than common variants. Moreover, whole-exome sequencing investigates variants that are not captured by SNP genotyping arrays. Wu et al.\textsuperscript{67,68} performed a systematic review of genetic research on schizophrenia and set up an open database (http://www.szdb.org) containing all of the identified findings from exome sequencing publications from inception to May 2020. Of the 16 studies, 15 evaluated essentially Caucasian samples and one assessed Asian subjects. Although none of the studies were conducted in Latin America, two were performed in the United States and included Latino samples, representing 6\textsuperscript{66} and 1\textsuperscript{70} of all subjects.

Altogether, this comprehensive review demonstrates that few genetic studies with large samples have been
carried out in Latin America, which, consequently, hinders our understanding of ethnic diversity and allele frequency in psychotic disorders. As such, international collaborations within Latin America represent a possible solution for these issues.

**Description of ongoing projects and the results so far**

Finally, in addition to published papers, we searched among leading regional schizophrenia researchers for current initiatives aimed at building large and thoroughly assessed cohorts to change this scenario. The Neuropsychiatric Genetics of Psychosis in Mexican Populations (NeuroMEX), Paisa, and Latin American Network for the Study of Early Psychosis (ANDES) projects are the result of international collaborations between Latin America and high-income nations. These international consortia could be used as framework in the future, possibly encouraging further actions (additional information for each of the following studies can be found in Box 1, available as online-only supplementary material).

**The NeuroMEX Study**

NeuroMEX is a three-part collaboration between the Harvard T.H. Chan School of Public Health, the Stanley Center for Psychiatric Genetics at the Broad Institute of Massachusetts Institute of Technology and Harvard, and the National Institute of Psychiatry Ramón de la Fuente Muñiz (INPRFM) in Mexico City. The NeuroMEX Study aims to recruit a total of 8,000 participants throughout Mexico, including 4,000 psychosis cases and 4,000 controls in a large GWAS, combined with extensive phenotypic data related to psychosis, schizophrenia, and bipolar disorder. The study seeks to build capacity and train Mexican collaborators to ensure the advancement of neuropsychiatric genetics research in Mexico.

**Paisa Study**

Paisa encompasses several administrative regions in Colombia, including the departments of Antioquia, Caldas, Risaralda and Quindío. The largest cities in this region are Medellín, Manizales, and Pereira, and the whole area includes a population of approximately 9 million individuals. The Paisa population, which originated between the 15th and 19th century, primarily from the genetic admixture of Native American women and European men is considered a genetic isolate. The isolation provided by the Andes Mountains, the rapid demographic growth of 25 generations and the subsequent population bottleneck that occurred during the initial admixture resulted in patterns of allelic variation that differ considerably from neighboring populations. Parental ancestry analyses of this population have shown that the Y chromosome lineage consists of 94% European, 5% African, and 1% Native American; the mitochondrial DNA lineage is 90% Native American, 8% African, and 2% European. The reduced diversity of surnames introduced by the founders allows easily traceable genealogical antecedents. There is also an increased prevalence of autosomal recessive disorders in the region associated with high degrees of consanguinity. Such unique features make the Paisa population especially valuable for the study of rare variant associations.

Previous genetic investigations in the Paisa population have contributed to our understanding of several neuropsychiatric disorders: Alzheimer’s disease, attention deficit hyperactivity disorder, bipolar disorder, and schizophrenia. More recently, the University of California, Los Angeles and the Universidad de Antioquia in Colombia initiated a case control study in the Paisa region to evaluate the relationship between severe mental illness phenotypes and genetic variants that may contribute to disease risk and development. For this study, a cohort of 8,000 individuals with diagnoses of mood or psychotic disorders (bipolar disorder, major depression disorder and schizophrenia) and 2,000 healthy controls from the Paisa region are being recruited and systematically evaluated with extensive clinical and cognitive batteries, as well as with genome wide genomic analyses.

**ANDES Study**

ANDES is a consortium created in 2018 of 15 different groups from six Latin American countries: Argentina, Brazil, Bolivia, Chile, Colombia, and Mexico. Its main objective is to stimulate research collaborations within the region, aiming to capitalize on existing resources, facilitate knowledge transfer across groups, and focus on specific aspects of psychosis in these countries. In addition to genetic studies, the participating groups are also conducting clinical, cognitive, epidemiological, economic, and neuroimaging analyses. As a first step, though, the network has focused on harmonizing existing data.

To illustrate this point, a recent study included structural MRI images from 334 patients with schizophrenia and 262 controls and was able to examine the modulating effect of poverty and violence on the brain in the harsh Latin American environment. Currently, projects are analyzing previously acquired genetic data to examine the validity of PRS for Latin American populations.

**Discussion**

Although schizophrenia has a high heritability (around 80% based on twin studies), the largest available GWAS explained only one-quarter of the disease variance, reinforcing the need to explore additional genetic and environmental variants. But to conduct such investigations, it is imperative to increase the diversity of research populations.

Latin America can contribute to both, given the admixed genetics of Latin Americans and their intense setting. However, researchers must face tough challenges, particularly the scarcity of research funding in the region, which calls for strategies to promote innovative studies. A syndemic framework could benefit new data collection efforts by maximizing results and reducing costs, for example, through the assessment of correlated phenotypes, such as...
trauma, psychosis, and cardiovascular diseases, all enhanced by each other and the environment. Additionally, as we suggested, employing weighted poly-environmental risk scores across different socioeconomic contexts might provide some insight into its overall effect on psychotic disorders and how adverse exposures possibly catalyze genetic liability.

Regarding genetic investigations specifically, public datasets and mobile applications will become increasingly important tools to achieve deeper and cheaper phenotyping. But, again, the lack of financial support to develop this capacity results in problematic bottlenecks. In Colombia, for example, the PAISA project has been largely facilitated by the availability of nationwide electronic records, which are unfortunately not common across Latin America. Furthermore, the planned quick collection of large genetic samples should be paralleled with efforts to improve exposure phenotyping.69

The ongoing projects described herein represent critical initiatives to foster Latin America’s role in international research. The NeuroMEX and Paisa projects show how partnerships with high-income countries can accelerate processes and should be stimulated. Likewise, the ANDES network provides an example of how collaborations even among Latin American countries could increase samples and optimize resources.

In conclusion, Latin American countries share many characteristics, including poverty, economic inequality, and violence, but also differ substantially in terms of culture, language, and genetic backgrounds. As such, these countries work as a natural laboratory to understand pathways connecting genetic liability and environmental exposures in the trajectories of psychotic disorders and could help increase the ethnic diversity of the research. However, the continuing lack of funding support and overall encouragement of research represent a great challenge. Extensive funding and partnerships are, therefore, required not only to collect data, but also to build the capacity that will allow future research initiatives.

Acknowledgements

This study was financed in part by Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES; Finance Code 001). RAB has received grants from the following governmental research funding agencies: CAPES, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; 2016/022465 and 2011/50740-5).

Disclosure

RAB has received research grants from Janssen Cilag, Novartis, and Roche; he has been a forum consultant for Janssen, Novartis, and Roche; and has participated in speaker bureaus for Ache, Janssen, Lundbeck, and Novartis. AG has been a consultant and/or advisor to or has received honoraria from: Aché, Daichi-Sankyo, Torrent, Bayer, Cristalia, Daiichi-Sankyo, and Janssen. The authors report no conflicts of interest.

References

8. 23andMe. Research team [Internet]. 2020 [cited 2020 Jan 14]. research.23andme.com/research/


