Genetic basis of mental illness: Individual genetics and ‘racial’ ancestry impact mental illness, but most studies have focused only on white people

Mental illness is a growing public health problem. In 2019, an estimated 1 in 8 people around the world were affected by mental disorders like depression, schizophrenia or bipolar disorder. While scientists have long known that many of these disorders run in families, their genetic basis isn’t entirely clear. One reason why is that the majority of existing genetic data used in research is overwhelmingly from white people.

In 2003, the Human Genome Project generated the first “reference genome” of human DNA from a combination of samples donated by upstate New Yorkers, all of whom were of European ancestry. Researchers across many biomedical fields still use this reference genome in their work. But it doesn’t provide a complete picture of human genetics. Someone with a different genetic ancestry will have a number of variations in their DNA that aren’t captured by the reference sequence.

When most of the world’s ancestries are not represented in genomic data sets, studies won’t be able to provide a true representation of how diseases manifest across all of humanity. Despite this, ancestral diversity in genetic analyses hasn’t improved in the two decades since the Human Genome Project announced its first results. As of June 2021, over 80% of genetic studies have been conducted on people of European descent. Less than 2% have included people of African descent, even though these individuals have the most genetic variation of all human populations.
To uncover the genetic factors driving mental illness, I, Sinéad Chapman and our colleagues at the Broad Institute of MIT and Harvard have partnered with collaborators around the world to launch Stanley Global, an initiative that seeks to collect a more diverse range of genetic samples from beyond the U.S. and
Northern Europe, and train the next generation of researchers around the world. Not only does the genetic data lack diversity, but so do the tools and techniques scientists use to sequence and analyze human genomes. So we are implementing a new sequencing technology that addresses the inadequacies of previous approaches that don’t account for the genetic diversity of global populations.

Ethically and equitably expanding the diversity of genomics data can help improve care and reduce disparities.

Global partnerships for global data

To study the genetics of psychiatric conditions, researchers use data from genome-wide association studies that compare the genetic variations between people with and without a particular disease. However, these data sets are mostly based on people of European ancestry, largely because research infrastructure and funding for large-scale genetics studies, and the scientists conducting these studies, have historically been concentrated in Europe and the United States.

Follow the latest news and policy debates on sustainable agriculture, biomedicine, and other ‘disruptive’ innovations. Subscribe to our newsletter.

One way to close this gap is to sequence genetic data from diverse populations. My colleagues and I are working in close partnership with geneticists, statisticians and epidemiologists in 14 countries across four continents to study the DNA of tens of thousands of people of African, Asian and Latino ancestries who are affected by mental illness. We work together to recruit participants and collect DNA samples that are sequenced at the Broad Institute in Massachusetts and shared with all partners for analysis.
Prioritizing the voices and priorities of local communities and scientists is foundational to our work. All partners have joint ownership of the project, including decision-making and sample and data ownership and control. To do this, we build relationships and trust with the local communities we are studying and the local university leaders and scientists with whom we are partnering. We work to understand local cultures and practices, and adapt our collection methods to ensure study participants are comfortable. For example, because there are different cultural sensitivities around providing saliva and blood samples, we have adapted our practices by location to ensure study participants are comfortable.

We also freely share knowledge and materials with our partners. There is a two-way exchange of information between the Broad Institute and local teams on study progress and results, enabling continual learning, teaching and unity between teams. We strive to meet each other where we are by exchanging practices and training scientists to support the development of locally grown and locally led research programs.

Our collaboration with African research groups provides a prime example of our model. For example, our African research colleagues are co-leaders on the grants that fund the lab equipment, scientists and other staff for projects based at their study sites. And we help to support the next generation of African geneticists and bioinformaticians through a dedicated training program.

Analyzing variation
Collecting samples from more diverse populations is only half of the challenge.

Existing genomic sequencing and analysis technologies do not adequately capture genetic variation across populations from around the world. That’s because these technologies were designed to detect genetic variations based on reference DNA from people of European ancestry, and they reduce accuracy when analyzing sequences that aren’t derived from the reference genome. When these tools are applied to genetic data from other populations, they fail to detect much of the rich variation in their genomes. This can lead researchers to miss out on important biomedical discoveries.

To address this issue, we developed an approach to genome sequencing that can detect more genetic variation from populations around the world. It works by sequencing the exome – the less than 2% of the genome that codes for proteins – in high detail, as well as sequencing the 98% of the genome that does not code for proteins in less detail.

Different types of sequencing methods have pros and cons.

This combined approach reduces the trade-offs geneticists often have to make in sequencing projects. High-depth whole genome sequencing, which reads through the entire genome multiple times to get detailed data, is too costly to do on a large number of DNA samples. While low-coverage sequencing reduces costs by reading smaller segments of the genome, it may miss some important genetic variation. With our new technology, geneticists can get the best of both worlds: sequencing the exome in depth maximizes the likelihood of pinpointing specific genes that play a role in mental illness, while sequencing the whole genome less in depth allows researchers to process large numbers of whole genomes more cost-effectively.

**Personalizing medicine**

Our hope is that this new technology will allow researchers to sequence large sample sizes from a diverse range of ancestries to capture the full breadth of genetic variation. With a better understanding of the genetics of mental illness, clinicians and researchers will be better equipped to develop new treatments that work for everyone.

Genomic sequencing opened a new era of personalized medicine, which promises to deliver treatments tailored to each individual person. This can be done only if the genetic variations of all ancestries are represented in the data sets that researchers use to make new discoveries about disease and develop treatments.

**Hailiang Huang is an Assistant Professor of Medicine at Harvard University.**

A version of this article appeared originally at The Conversation and is posted here with permission. Check out The Conversation on Twitter @ConversationUS