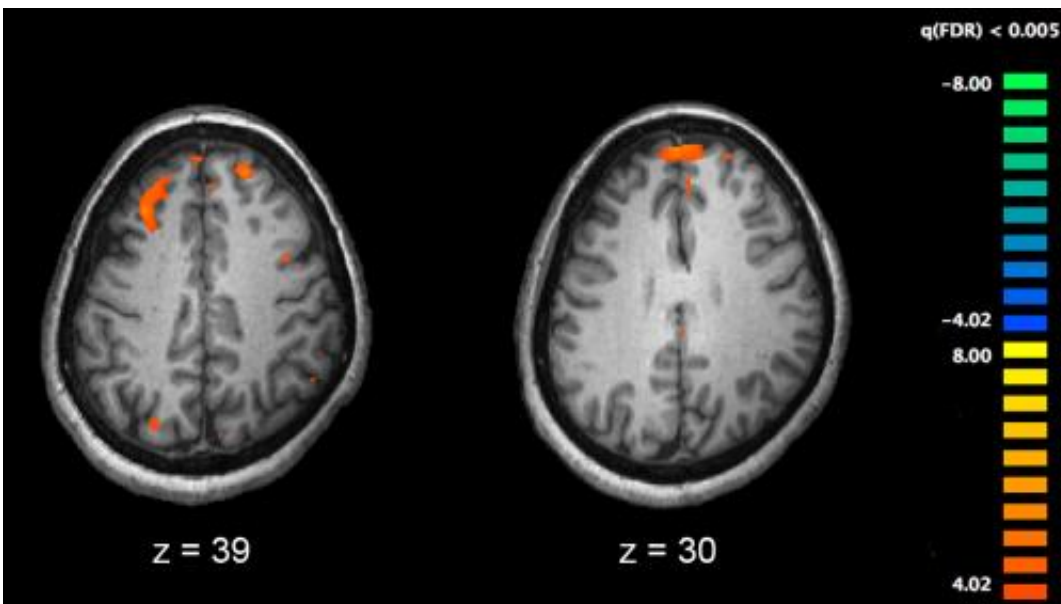# A machine learning approach for predicting risk of schizophrenia using a blood test

August 3 2021

Functional magnetic resonance imaging (fMRI) and other brain imaging technologies allow for the study of differences in brain activity in people diagnosed with schizophrenia. The image shows two levels of the brain, with areas that were more active in healthy controls than in schizophrenia patients shown in orange, during an fMRI study of working memory. Credit: Kim J, Matthews NL, Park S./PLoS One.

An innovative strategy that analyzes a region of the genome offers the possibility of early diagnosis of schizophrenia, reports a team led by researchers at Baylor College of Medicine. The strategy applied a machine learning algorithm called SPLS-DA to analyze specific regions

of the human genome called CoRSIVs, hoping to reveal epigenetic markers for the condition.

In DNA from blood samples, the team identified epigenetic markers, a profile of methyl chemical groups in the DNA, that differ between people diagnosed with schizophrenia and people without the disease and developed a model that would assess an individual's probability of having the condition. Testing the model on an independent dataset revealed that it can identify schizophrenia patients with 80% accuracy. The study appears in the journal *Translational Psychiatry*.

"Schizophrenia is a devastating disease that affects about 1% of the world's population," said corresponding author Dr. Robert A. Waterland professor of pediatrics – nutrition at the USDA/ARS Children's Nutrition Research Center at Baylor and of molecular and human genetics. "Although genetic and environmental components seem to be involved in the condition, current evidence only explains a small portion of cases, suggesting that other factors, such as epigenetic, also could be important."

Epigenetics is a system for molecular marking of DNA—it tells the different cells in the body which genes to turn on or off in that cell type, therefore epigenetic markers can vary between different normal tissues within one individual. This makes it challenging to assess whether epigenetic changes contribute to diseases involving the brain, like schizophrenia.

To address this obstacle, Waterland and his colleagues had identified in previous work a set of specific genomic regions in which DNA methylation, a common epigenetic marker, differs between people but is consistent across different tissues in one person. They called these genomic regions CoRSIVs for correlated regions of systemic interindividual variation. They proposed that studying CoRSIVs is a

novel way to uncover epigenetic causes of disease.

"Because methylation patterns in CoRSIVs are the same in all the tissues of one individual, we can analyze them in a blood sample to infer epigenetic regulation on other parts of the body that are difficult to assess, such as the brain," Waterland said.

Many previous studies have analyzed methylation profiles in blood samples with the goal of identifying epigenetic differences between individuals with schizophrenia, the researchers explained.

"Our study is innovative in various ways," said first author Dr. Chathura J. Gunasekara, computer scientist in the Waterland lab. "We focused on CoRSIVs and also applied for the first time the SPLS-DA machine learning algorithm to analyze DNA methylation. As a scientist interested in applying machine learning to medicine, our findings are very exciting. They not only suggest the possibility of predicting risk of schizophrenia early in life, but also outline a new approach that may be applicable to other diseases."

The current study also is innovative because it considered major potential confounding factors other studies did not take into account. For instance, methylation patterns in blood can be affected by factors such as smoking and taking antipsychotic medications, both of which are common in schizophrenia patients.

"Here, we took various approaches to evaluate whether the [methylation patterns](link) we detected at CoRSIVs were affected by medication use and smoking. We were able to rule that out," Waterland said. "This, together with the fact that DNA methylation at CoRSIVs is established very early in life, indicates that the epigenetic differences we identified between schizophrenia patients and healthy individuals were there before the disease was diagnosed, suggesting they may contribute to the condition."

Using this novel approach, the researchers were able to achieve much stronger epigenetic signals associated with schizophrenia than has ever been done before, said the team.

"We consider our study a proof of principle that focusing on CoRSIVs makes epigenetic epidemiology possible," Waterland said.

More information: Chathura J. Gunasekara et al, A machine learning case–control classifier for schizophrenia based on DNA methylation in blood, *Translational Psychiatry* (2021). DOI: 10.1038/s41398-021-01496-3

Provided by Baylor College of Medicine