SCIENCE

A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG **MAY 17, 2019**

SEAN NEL / SHUTTERSTOCK

In 1996, a group of European researchers found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even predicting a person's response to

antidepressants.

Back then, tools for sequencing DNA weren't as cheap or powerful as they are today. When researchers wanted to work out which genes might affect a disease or trait, they made educated guesses, and picked likely "candidate genes." For depression, *SLC6A4* seemed like a great candidate: It's responsible for getting a chemical called serotonin into brain cells, and serotonin had already been linked to mood and depression. Over two decades, this one gene inspired at least 450 research papers.

But a new study—the biggest and most comprehensive of its kind yet—shows that this seemingly sturdy mountain of research is actually a house of cards, built on nonexistent foundations.

Richard Border of the University of Colorado at Boulder and his colleagues picked the 18 candidate genes that have been most commonly linked to depression—*SLC6A4* chief among them. Using data from large groups of volunteers, ranging from 62,000 to 443,000 people, the team checked whether any versions of these genes were more common among people with depression. "We didn't find a smidge of evidence," says Matthew Keller, who led the project.

Between them, these 18 genes have been the subject of more than 1,000 research papers, on depression alone. And for what? If the new study is right, these genes *have nothing to do with depression*. "This should be a real cautionary tale," Keller adds. "How on Earth could we have spent 20 years and hundreds of millions of dollars studying pure noise?"

"What bothers me isn't just that people said [the gene] mattered and it didn't," wrote the pseudonymous blogger Scott Alexander in a widely shared post. "It's that we built whole imaginary edifices on top of this idea of [it] mattering." Researchers studied how *SLC6A4* affects emotion centers in the brain, how its influence varies in different countries and demographics, and how it interacts with other genes. It's as if they'd

been "describing the life cycle of unicorns, what unicorns eat, all the different subspecies of unicorn, which cuts of unicorn meat are tastiest, and a blow-by-blow account of a wrestling match between unicorns and Bigfoot," Alexander wrote.

[Read: Why are so many teen athletes struggling with depression?]

Border and Keller's study may be "bigger and better" than its predecessors, but "the results are not a surprise," says Cathryn Lewis, a geneticist at Kings College London. Warnings about the *SLC6A4*/depression link have been sounded for years. When geneticists finally gained the power to cost-efficiently analyze entire genomes, they realized that most disorders and diseases are influenced by thousands of genes, each of which has a tiny effect. To reliably detect these minuscule effects, you need to compare hundreds of thousands of volunteers. By contrast, the candidate-gene studies of the 2000s looked at an average of *345 people*! They couldn't possibly have found effects as large as they did, using samples as small as they had. Those results must have been flukes—mirages produced by a lack of statistical power. That's true for candidate-gene studies in many diseases, but Lewis says that other researchers "have moved on faster than we have in depression."

Marcus Munafò of the University of Bristol remembers being impressed by the early *SLC6A4* research. "It all seemed to fit together," he says, "but when I started doing my own studies in this area, I began to realize how fragile the evidence was." Sometimes the gene was linked to depression; sometimes it wasn't. And crucially, the better the methods, the less likely he was to see such a link. When he and others finally did a large study in 2005—with 100,000 people rather than the 1,000 from the original 1996 paper—they got nothing.

"You would have thought that would have dampened enthusiasm for that particular candidate gene, but not at all," he says. "Any evidence that the results might not be reliable was simply not what many people wanted to hear." In fact, the pace at which *SLC6A4*/depression papers

were published *accelerated* after 2005, and the total number of such papers quadrupled over the next decade. "We're told that science self-corrects, but what the candidate-gene literature demonstrates is that it often self-corrects very slowly, and very wastefully, even when the writing has been on the wall for a very long time," Munafò adds.

Many fields of science, from psychology to cancer biology, have been dealing with similar problems: Entire lines of research may be based on faulty results. The reasons for this so-called reproducibility crisis are manifold. Sometimes, researchers futz with their data until they get something interesting, or retrofit their questions to match their answers. Other times, they selectively publish positive results while sweeping negative ones under the rug, creating a false impression of building evidence.

Beyond a few cases of outright misconduct, these practices are rarely done to deceive. They're an almost inevitable product of an academic world that rewards scientists, above all else, for publishing papers in high-profile journals—journals that prefer flashy studies that make new discoveries over duller ones that check existing work. People are rewarded for being *productive* rather than being *right*, for building ever upward instead of checking the foundations. These incentives allow weak studies to be published. And once enough have amassed, they create a collective perception of strength that can be hard to pierce.

Terrie Moffitt of Duke University, who did early influential work on *SLC6A4*, notes that the candidate-gene approach has already been superseded by other methods. "The relative volume of candidate-gene studies is going way down, and is highly likely to be trivial indeed," she says. Border and Keller disagree. Yes, they say, their geneticist colleagues have largely abandoned the approach, which is often seen as something of a historical embarrassment. "But we have colleagues in other sciences who had no idea that there was even any question about these genes, and are doing this research to this day," Border says. "There's not good communication between subfields." (A few studies on *SLC6A4* and

depression have even emerged since their study was published in March.)

[*Read: The link between Neanderthal DNA and depression risk*]

The goalposts can also change. In one particularly influential study from 2003, Avshalom Caspi, Moffitt, and others claimed that people with certain versions of *SLC6A4* were more likely to become depressed after experiencing stressful life events. Their paper, which has been cited more than 8,000 times, suggested that these genes have subtler influences, which only manifest in certain environments. And if bigger studies found that the genes had *no* influence, it's probably because they weren't accounting for the experiences of their volunteers.

Border and Keller have heard that argument before. So, in their study, they measured depression in many ways—diagnosis, severity, symptom count, episode count—and they accounted for environmental factors such as childhood trauma, adulthood trauma, and socioeconomic adversity. It didn't matter. No candidate gene influenced depression risk in any environment.

But Suzanne Vrshek-Schallhorn of the University of North Carolina at Greensboro says that Border's team didn't assess life experiences with enough precision. "I cannot emphasize enough how insufficient the measures of the environment used in this investigation were," she says. "Even for measures that fall below gold-standard stress-assessment approaches, they represent a new low." By using overly simple yes-or-no questionnaires rather than more thorough interviews, the team may have completely obscured any relationships between genes and environments, Vrshek-Schallhorn claims. "We should not get starry-eyed about large sample sizes, when measure validity is compromised to achieve them. We need to emphasize both quality and quantity."

But Border argues that even if there had been "catastrophic measurement error," his results would stand. In simulations, even when

he replaced half the depression diagnoses and half the records of personal trauma with *coin flips*, the study would have been large enough to detect the kinds of effects seen in the early candidate-gene papers.

Similar debates have played out in other fields. When one group of psychologists started trying to reproduce classic results in much larger studies, their peers argued that any failures might simply be due to differences between the new groups of volunteers and the originals. This excuse has eroded with time, but to Border, it feels familiar. "There's an unwillingness to part with a previous hypothesis," he says. "It's hard to wrap your head around the fact that maybe you were on a wild goose chase for years."

Keller worries that these problems will be used as ammunition to distrust science as a whole. "People ask, *Well, if scientists are publishing crap, why should we believe global warming and evolution?*" he says. "But there's a real difference: Some people were skeptical about candidate genes even back in the 1990s. There was never unanimity or consensus in the way there is for human-made global warming and the theory of evolution."

Nor, he says, should his work be taken to mean that genes don't affect depression. They do, and with newer, bigger studies, researchers are finally working out *which* ones do. If anything, the sordid history of the candidate-gene approach propelled the development of better methods. "I feel like the field of psychiatric genetics felt really burned coming out of the candidate-gene era, and took strides to make sure it won't happen again." That includes sharing data openly, and setting standards for how large and powerful studies need to be.

Dorothy Bishop of the University of Oxford argues that institutions and funders that supported candidate-gene work in depression should also be asking themselves some hard questions. "They need to recognize that even those who think they are elite are not immune to poor reproducibility, which leads to a huge amount of waste," she says.

"We have got to set up a system, or develop a culture, that rewards people for actually trying to do it right," adds Keller. "Those who don't learn from the past are doomed to repeat it."

We want to hear what you think about this article. [Submit a letter](#) to the editor or write to letters@theatlantic.com.